PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12N 15/53, 9/04, 1/21 | A1

(11) International Publication Number: **WO 98/51802**

(43) International Publication Date: 19 November 1998 (19.11.98)

(21) International Application Number: PCT/US98/09627

(22) International Filing Date: 12 May 1998 (12.05.98)

(30) Priority Data:
60/046,182 12 May 1997 (12.05.97) US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application
US 60/046,182 (CIP)
Filed on 12 May 1997 (12.05.97)

(71) Applicant *(for all designated States except US)*: THERMOGEN, INC. [US/US]; 2201 W. Campbell Park Drive, Chicago, IL 60612 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: DEMIRJIAN, David, C. [US/US]; Chicago, IL 60615 (US). BRIKUN, Igor, A. [US/US]; Forest Park, IL 60130 (US). CASADABAN, Malcolm, J. [US/US]; Chicago, IL 60615 (US). VONSTEIN, Veronika [US/US]; Chicago, IL 60615 (US).

(74) Agent: CHAO, Mark; McDonnell Boehnen Hulbert & Berghoff, 300 South Wacker Drive, Chicago, IL 60606 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: METHOD FOR THE STABILIZATION OF PROTEINS AND THE THERMOSTABILIZED ALCOHOL DEHYDROGENASES PRODUCED THEREBY

(57) Abstract

The present invention provides a method for the directed evolution of proteins, particularly a method for improving the thermostability of proteins, particularly alcohol dehydrogenases, and especially horse liver alcohol dehydrogenase. The present invention also provides thermostabilized alcohol dehydrogenases produced according to this method.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

METHOD FOR THE STABILIZATION OF PROTEINS AND THE THERMOSTABILIZED ALCOHOL DEHYDROGENASES PRODUCED THEREBY

TECHNICAL FIELD OF THE INVENTION

The present invention generally relates to a method for the directed evolution of proteins. In particular, the method is directed to stabilization of proteins such as dehydrogenases, and particularly is directed to a method for improving the thermostability of dehydrogenases such as alcohol dehydrogenases. The present invention also relates to thermostabilized alcohol dehydrogenases produced according to this method.

BACKGROUND OF THE INVENTION

Biocatalysts are enzymes which can specifically and efficiently expedite chemical reactions such as the synthesis of chemical compounds and biopolymers (Dixon et al., Enzymes (Academic Press, New York: 1979)). Biocatalysts are the key players in a number of important industrial synthetic and degradative applications including, but not limited to, the following:

- *Synthetic Applications* - Biocatalysts currently are employed as feasible alternatives to traditional catalysts, especially for the synthesis of chiral intermediates, or in the reduction of the number of protection/deprotection steps.
- *Biodegradation Applications* - Biocatalysts currently are employed as enzymatic degradation agents for environmental pollutants such as PCBs, chlorinated hydrocarbons, RDX, halogenated organic compounds, TNT, and other byproducts of industrial production that present significant health risks.
- *Diagnostics and Biosensors* - Biocatalysts currently are employed as detection agents in diagnostic tests and as biosensors which require enzyme durability.
- *Other large-scale industrial applications* - Biocatalysts currently are employed as catalysts in

the production of fuel supplies through conversion of agricultural feedstocks.

One enzyme that is of considerable utility in current enzymatic processes is the dehydrogenase. In particular, alcohol dehydrogenases are enzymes that command formal, reversible, two-electron chemistry in which alcohols are oxidized to the corresponding ketones. Depending on the precise reaction conditions, ketones can be reduced to the respective alcohols via a stereospecific delivery of a hydride equivalent catalyzed by the enzyme coupled to a bound cofactor such as NADH or NADPH (Lemiere, "Alcohol Dehydrogenase Catalyzed Oxidoreduction Reactions in Organic Chemistry", In Enzymes as Catalysts in Organic Synthesis, Schneider et al., Eds. (1986) p. 17). This system thus provides a mild, extremely sensitive route to chiral compounds, without contamination from undesired, competing reactions.

Such chiral compounds can be used, especially by the pharmaceutical industry, for the preparation of chiral therapeutics, and for effectively generating a wide variety of compounds having the capacity for industrial scale-up (Seebach et al., Org. Synth., 63, 1-_ (1984); Bradshaw et al., J. Org. Chem., 57, 1532(1992); Hummel, Biotechnol. Lett., 12, 403(1990)). In particular, dehydrogenases show promise for commercial application in the preparation of unusual amino acids and β-hydroxyketones, and in the resolution of racemic alcohols (Benoiton et al., J. Am. Chem. Soc., 79, 6192 (1957); Casy et al., Tetrahedron Lett., 33, 817 (1992); Jacovac et al., J. Am. Chem. Soc., 104, 4659-4665 (1982); Jones et al. Can. J. Chem., 60, 19 (1982)). Of the dehydrogenases, horse liver alcohol dehydrogenase (HLADH) is one of the most commonly used.

For an enzyme biocatalyst such as HLADH to prove useful in a wide-scale, practical, industrial

application, it is important that the biocatalyst possess the ability to survive harsh, dynamic, environmental and handling conditions inherent to large-scale commercial processes. These conditions include nonrefrigerated storage, and exposure to organic cosolvents and high reaction temperatures, as well as more idiosyncratic demands imposed by a particular industrial application.

To date, one of the greatest challenges associated with biocatalyst implementation is that of overcoming an overall intrinsic instability that results in a requirement for special preparative approaches and handling conditions. Many methods have been used in an attempt to stabilize certain proteins. Rational protein engineering has allowed the redesign of proteins with altered properties such as enhanced stability, shifted pH optima, and different substrate specificities (see, e.g., Bryan et al., Proteins, 1, 326-334 (1986); Pantoliano et al., Biochemistry, 26, 2077-82 (1987); Carter et al., Science, 237, 394-399 (1987); Wells et al., "Designing substrate specificity by protein engineering of electrostatic interactions", __, 84,1219-1223(1987); Grutter et al., Nature, 277, 667-669 (1979)).

While potentially an extremely powerful tool, rational protein engineering can be extremely time-consuming and expensive, and currently can be employed only for a very small number of enzymes having well-defined crystal or solution structures. Moreover, since the approach is tailored to a specific enzyme, it typically cannot be generalized to other enzyme species. Other post-production stabilization methods such as immobilization (Macaskie et al., FEMS Microbiol Rev., 14,351-67 (1994); Shtelzer et al., Biotechnol. Appl. Biochem., 15, 227-35 (1992); Phadke, Biosystems, 27, 203-6 (1992)), or use of cross-linked enzymes (Navia et al., "Crosslinked enzyme crystals as robust biocatalysts", Proceedings of the Materials Research Society 1993 Symposium, Biomolecular Materials by Design (1993)),

suffer some of the same as well as further shortcomings, and similarly, are often too expensive to implement.

By contrast, directed evolution potentially can provide a practical approach to tailoring enzymes for a wide range of applications (Shao et al., "Engineering New Functions and Altering Existing Functions", Current Opinion in Structural Biology, *in press* (1996)). In support of this, enzymes have been shown to be highly adaptable molecules over evolutionary time scales. Many enzymes catalyzing very different reactions appear to have come about by divergent evolution, acquiring diverse capabilities by the processes of random mutation, recombination, and natural selection.

Thus, there remains a need for an effective means to randomly engineer better enzymes, particularly dehydrogenases, and especially, HLADH. The present invention seeks to overcome some of the aforesaid problems of enzyme design. In particular, it is an object of the present invention to provide a method for the directed evolution of enzymes, particularly dehydrogenases, and especially HLADH. It further is an object of the present invention to provide a method for stabilizing, e.g. improving the thermostability of enzymes such as dehydrogenases. Such a method of stabilizing dehydrogenases (particularly HLADH) would present a major advancement in the field since it would extend the shelf life, longevity, and active temperature range of these enzymes. These and other objects and advantages of the present invention, as well as further inventive features, will be apparent from the description of the invention provided herein.

BRIEF SUMMARY OF THE INVENTION

Briefly, the present invention provides, *inter alia*, a method for the stabilization of a protein (particularly for the stabilization of an alcohol dehydrogenase such as horse liver alcohol dehydrogenase (HLADH), general

enrichment/selection means that can be employed in *Escherichia* and *Thermus* to select for cells having altered levels of alcohol dehydrogenase activity as compared to a wild-type cell, thermostabilized HLADH proteins and nucleic acid sequences encoding same, as well as plasmids and hosts cells comprising the nucleic acid sequences.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 is a diagram that generally depicts the approach of the present invention for the accelerated evolution of enzymes. A pool of mutants of the particular gene is obtained by means such as spontaneous, directed, chemical, or PCR-mediated mutagenesis. The mutants of interest (i.e., having the particular stabilized feature) are identified by means of a screen or selection (**A**), and optionally, compatible mutations can be combined (e.g., by gene splicing, *in vitro* recombination, and the like) to enhance the stability even further (**B**).

Figure 2 is a digitized image of results of a filter assay for alcohol dehydrogenase activity which demonstrates that wild-type HLADH is rapidly inactivated at 75°C: no heat treatment (**A**); 5 minutes of heat treatment at 75°C (**B**); 10 minutes of heat treatment at 75°C (**C**); 15 minutes of heat treatment at 75°C (**D**); 20 minutes of heat treatment at 75°C (**E**); and 50 minutes of heat treatment at 75°C (**F**).

Figure 3 is a partial restriction map of the plasmid pTG450 which contains the *adh* gene from plasmid pBPP cloned into a pTG100kan^{tr2} *Thermus* shuttle vector.

Figure 4 is a bar chart that depicts the increased thermostability of HLADH mutants produced according to the invention at 70°C. Cells containing pGEM-T (i.e., having no HLADH gene) did not show any HLADH activity.

Figure 5 is the sequence of *adh* gene [SEQ ID NO:1] that encodes the HLADH protein [SEQ ID NO:2], with the

location of certain mutations produced according to the invention identified as the boxed regions.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides, among other things, a method for stabilizing a certain feature of a protein (e.g., stability at a certain temperature, stability in the presence of certain reagents, etc.). In particular, the method of the invention provides a method for thermostabilizing a protein. Namely, the invention preferably provides a method of obtaining nonnative protein having a thermostability that is increased over that of the native version of said protein, as further described herein.

According to the invention, a "native" protein is the protein as it generally is found in nature. By contrast, a "nonnative" protein differs from the native protein in that it has been modified by human intervention, i.e., at either the level of the protein or its encoding DNA (e.g., by recombinant means to directly alter the genome; by unique selection and forced mutation; by random mutagenesis). Moreover, a "protein" desirably can be either an entire protein, or a portion of a protein (e.g., as where a chimeric nonnative protein results from either transcriptional or translational gene fusion). Similarly, a "nonnative protein" in some applications (e.g., applications for further study) may be a peptide (i.e., an incomplete protein), as where the peptide is chemically synthesized or, where a gene's coding sequence is transcribed or translated *in vitro* or, is produced by chemical processing of a complete protein.

A preferred protein for stabilization, particularly thermostabilization according to the invention is a dehydrogenase, particularly an alcohol dehydrogenase, and especially horse liver alcohol dehydrogenase (e.g., as obtained from plasmid pBPP, and/or as set forth in

SEQ ID NO:2). Notably, with respect to SEQ ID NO:2, this protein does not initiate with methionine (Met). However, other varients of horse liver alcohol dehydrogenase produced by *in vitro* synthetic reactions, by means of chemical synthesis or, in other hosts (e.g., an eukaryotic host or other prokaryotic host cell) may possess a methionine residue in the first position of the protein. The numbering of residues in such proteins of course, would differ somewhat from that of SEQ ID NO:2. Namely, the second position of the aforementioned protein would be equivalent to the first position of the protein of SEQ ID NO:2. Of course, the ordinarily skilled artisan would know how to compare equivalent regions of proteins.

Desirably, other proteins (particularly proteins having capacity for industrial implementation) can be stabilized (e.g., thermostabilized) according to the invention. For instance, an alcohol dehydrogenase protein can be employed from another species. It is anticipated that this approach can be employed with alcohol dehydrogenases from other species based on the similarities between certain of the various alcohol dehydrogenases. Also, a protein according to the invention optionally can be another type of dehydrogenase, e.g., another type of NAD+(P)-linked dehydrogenase including, but not limited to, malate dehydrogenase, lactate dehydrogenase, isocitrate dehydrogenase (NADP+), hydroxylacyl CoA dehydrogenase, glyceraldehyde 3-phosphate dehydrogenase, and glucose 6-phosphate dehydrogenase (NADP+).

In a preferred embodiment, the method can be employed to thermostabilize a horse liver alcohol dehydrogenase. This method generally is depicted in Figure 1. Preferably the method comprises:

(a) obtaining in a vector a gene that encodes the native protein;

(b) mutating the vector at more than one position in the gene to produce a vector library of cells comprising mutated versions of the gene;

(c) introducing the vector library *en masse* into cells of a strain in which the majority of the mutated versions of the gene are transcribed and translated to produce a cell library;

(d) screening the cell library to identify a cell comprising a mutated version of the gene that encodes a nonnative protein having a thermostability that is increased over that of the wild-type verson of the protein; and

(e) purifying the cell from the cell library.

According to the invention, "gene that encodes said protein" can comprise a recombinant or nonrecombinant sequence, i.e., a sequence that is present as found in nature (i.e., encodes a native amino acid sequence) or, has been modified, for instance by the introduction of mutations (e.g., point mutations, insertions, deletions, or rearrangements) to comprise a nonnative amino acid sequence or, can be a mixture of native and nonnative amino acid sequences. Similarly, a recombinant gene may conjoin coding sequences (either in entirety or in part) with regulatory sequences (e.g., transcription initiation, transcription termination, translational start or stop sites, protein secretion sequences, and the like) which are not typically conjoined in nature. This can allow the production of a protein in a host in which it normally is not produced (e.g., production of a eukaryotic protein in a prokaryotic cell). Preferably, however, the recombinant gene (which can derive, in entirety or part, from any prokaryotic, eukaryotic, bacteriophage, or viral source) is capable of being transcribed and translated in a prokaryotic cell, particularly, a cell comprising a member of the genuses *Escherichi* or *Thermus*.

Thus, preferably a host cell in the context of the present invention (i.e., which can be employed in a method of stabilizing proteins) is a member of the kingdom Bacteria, Archaea, or Eukarya. In particular, preferably a cell employed in the method of stabilizing (particularly thermostabilizing) proteins according to the invention is a thermophile or hyperthermophile. In particular, preferably a cell is a member of the genus *Thermus*, and desirably is of the species *Thermus flavus*, *Thermus aquaticus*, *Thermus thermophilus*, or *Thermus sp.* Optimally a cell is either an *Escherichia coli* cell or a *Thermus aquaticus* cell.

The vector in which the gene of interest is subcloned can be any vector appropriate for delivery of a gene to a cell. For instance, the vector can be a plasmid, bacteriophage, virus, phagemid, cointegrate of one or more vector species, etc. Optimally, however, a vector is one that can be employed for gene expression in a prokaryotic cell such as a *Thermus* or *Eshcerichia* cell. It also is preferable that a vector have an ability to shuttle between different cells, e.g., between a *Thermus* and an *Eschericia* cell. One such vector that can be employed in the context of the invention is the vector pTG450.

The preferred method of the invention calls for mutating a vector containing the gene encoding the protein to be stabilized. Any method of mutagenesis such as is known to those skilled in the art and particularly as is described in the following Examples, can be employed in the method of the invention for generating a mutated gene. Desirably a PCR-based (error prone) approach, especially as set out as follows, is employed for mutagenesis. However, other mutagens (e.g., chemical mutagens such as hydroxylamine), also can be employed.

In the preferred method of mutagenesis employed in the invention, desirably the vector is mutated at more than one position in the gene of interest. This can be

assessed by means known in the art and as described in the Examples. Such mutagenesis in more than one position in the gene will result in a "vector library" comprising mutated versions of a gene, particularly of a horse liver alcohol dehydrogenase gene, which are present in the library mixture.

The vector library can be introduced *en masse* into cells (e.g., by transformation). Since the vectors and the cells employed for these methods are selected to be compatible, and the gene is engineered (e.g., as described below) to contain or to be flanked by any sequences necessary for its expression, it is expected that such introduction will result in the transcription and ensuing translation of the introduced gene. Moreover, such *en masse* introduction will result in the generation of a cell library comprising a mixture of cells transformed with plasmids having differing mutated genes. In some instances, it may be desirable to reisolate the vectors from the cell library (e.g., by a plasmid isolation or other vector isolation protocol), excise out the mutated gene, and subclone the mutated gene into another vector (e.g., a vector that has not been mutagenized).

Following the generation of the cell library, the cells preferably are screened under conditions that allow identification of a cell comprising a mutated version of the gene of interest that encodes a nonnative protein having a protein that is stabilized (e.g., thermostabilized) over that of the wild-type (i.e., native) versions of the protein. A variety of selection means can be employed in accordance with the method of the present invention and, in particular, the selection means identified in the Examples which follow can be employed. Of course, one of ordinary skill in the art could modify these methods such that they are adapted for a particular host cell and/or a particular protein of interest. Desirably, however, screening conditions

are employed that provide for enrichment and/or selection for a cell containing nonnative DNA that encodes a protein having a particular feature of interest.

In particular, when the protein being stabilized according to the invention is an alcohol dehydrogenase, and particularly HLADH, the screen preferably can be carried out at increased temperature. For instance, desirably, screening is done at temperature a few degrees above and a few degrees below the temperature at which the native (i.e., wild-type) alcohol dehydrogenase is inactivated in the particular host cell employed for screening.

According to this invention, "increasing the thermostability" of a nonnative protein means: (a) increasing the length of time at which a nonnative protein exhibits activity as compared to the wild-type protein; (b) increasing the temperature at which a nonnative protein exhibits activity as compared to a wild-type protein; or (c) increasing the length of time and temperature at which a nonnative protein exhibits activity as compared to a wild-type protein. A protein's activity can be determined by a variety of tests that differ with the various proteins to be tested. A few representative tests that can be employed in the method of the invention are set out in the following Examples. Preferably, however, "activity" means a detectable activity ranging from 10 to 90 units. For instance, whereas a wild-type protein might exhibit 10% activity at a defined temperature for a set amount of time, a thermostabilized enzyme might exhibit 10% activity at the same temperature for an increased amount of time, and/or might exhibit an activity at an increased temperature at which the native protein exhibits reduced or no activity.

The screening methods also desirably can be done, for instance, in the presence of alcohol, optionally at a lowered pH.

Following screening of cells to identify those having the desired trait(s) imparted by the mutated gene, optionally, cells exhibiting the trait can be further isolated. Vectors containing mutated versions of the gene of interest optionally can be further mutagenized by repeating steps (b) through (e) above to further stabilize the encoded protein.

The present invention accordingly also provides screens that can be employed to select for or against cells having altered ADH activity. For instance, the invention provides a method for selecting against growth of *Eschericia coli* recombinant cells which comprise levels of alcohol dehydrogenase that are higher than those of wild-type *Eschericia coli* cells. According to this invention, "growth" means an increase in cell mass, or some other evidence of cell metabolism such as one of ordinary skill in the art knows how to detect, or is described in the following Examples. An "absence of growth" means growth is not measurable by common procedures (e.g., visual or spectrophotometric observation and the like) or, cell killing. Cell killing can be determined by any well known means, e.g., visual observation, release of cell components, vital staining etc.

Thus the E.coli selection method comprises growing said recombinant cells under conditions selected from the group consisting of, wherein ethanol is present in a concentration of about 10%, isopropanol is present in a concentration of about 4%, and propanol is present in a concentration of about 2%, with the proviso that the wild-type cells exhibit reduced or an absence of growth under these conditions.

The present invention similarly provides a method for selecting for growth of *Thermus flavus* recombinant

cells which comprise levels of alcohol dehydrogenase that are higher than those of wild-type *Thermus flavus* cells. This method comprises growing the recombinant cells under conditions selected from the group consisting of wherein ethanol is present at a concentration of about1% in a liquid or solid medium at a pH of about 7.0, with the proviso that the wild-type cells exhibit reduced or an absence of growth under these conditions.

As mentioned previously, these methods have been employed to thermostabilize HLADH. In particular, the invention provides an isolated and purified thermostabilized HLADH protein comprising a sequence selected from the group consisting of SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18 and SEQ ID NO:20. The invention also provides genes encoding such protein, e.g., an isolated and purified nucleic acid comprising a sequence selected from the group consisting of SEQ ID NO:3; SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9, SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17 and SEQ ID NO:19.

Moreover, the invention provides for plasmids encoding for such proteins: e.g., a plasmid comprising one of the aforementioned nucleic acid sequences; and a plasmid selected from the group consisting of pAD7; pAD8, pAD10, pAD91, pAD92, pAD93, pAD95, pAD111, pAD113, and pTG450.

The invention further preferably provides a method of increasing the thermostability of horse liver alcohol dehydrogenase. This method comprises introducing into a gene which encodes the alcohol dehydrogenase a mutation at a codon which codes for an amino acid residue at a position selected from the group consisting of the amino acid positions, 75, 94, 110, 177, 257, 268, 282, 292, and 297.

Examination of the three-dimensional structure of the HLADH protein will elucidate the manner in which further amino acid substitutions thermostabilizing the enzyme can be made, for instance, like-for-like (e.g., with acidic amino acids (i.e., aspartic acid, glutamic acid) being substituted for acidic amino acids; basic amino acids (i.e., lysine, arginine, histidine) being substituted for basic amino acids; sulfur containing amino acids (i.e., cysteine) being substituted for sulfur containing amino acids; amides (i.e., asparagine, glutamine) being substituted for amides, aliphatic nonpolar amino acids (i.e., glycine, alanine, valine, leucine, isoleucine) being substituted for aliphatic nonpolar amino acids; and alcoholic, aliphatic, and aromatic amino acids (i.e., serine, threonine, thyrosine, phenylalanine, and tryptophan) being substituted for alcoholic, aliphatic, and aromatic amino acids.

Additional uses and benefits of the invention will be apparent to one of ordinary skill in the art.

EXAMPLES

The following examples further illustrate the present invention but, of course, should not be construed as in any way limiting its scope.

EXAMPLE 1: Quantitative assay for ADH in cell extracts.

This example describes a method for the quantification of ADH in cell extracts, particularly for the quantitation of HLADH, that can be used according to the invention.

For this assay, overnight cultures of cells to be assayed are grown in rich media. The cells are washed, resuspended in 600 µl of assay buffer (83 mM KH_2PO_4 [pH 7.3], 40 mM KCl, 0.25 mM EDTA), and sonicated. The assay

mixture contains 500 µl of cell extract, 100 µl EtOH, 20 µl 100 mM NAD, 830 µl buffer and is carried out at room temperature. The reaction is run for 3 minutes and absorbence at 340 nM is measured. Using this approach it is possible to identify a high IPTG inducible activity in the strains with the HLADH coding sequence under the control of the *lac*Z promoter. This method thus produces a reliable quantitative determination of HLADH activity present in the cell.

EXAMPLE 2: *p*-Rosanaline/alcohol plate screen in *E. coli*.

This example describes a plate screen for ADH activity that can be employed, for instance, in *E. coli*.

p-Rosaniline indicator plates are prepared according to Conway et al. (Conway et al., 169, 2591-2597 (1987)) by adding 8 ml of *p*-rosaniline (2.5 mg/ml in 96% ethanol) and 100 mg of sodium bisulfite to 400 ml batches of precooled (45°C) Luria agar. Most of the dye is immediately converted to the leuco form by reaction with bisulfite to produce a rose-colored medium. Ethanol diffuses into the *E. coli* cells to produce the acetaldehyde by alcohol dehydrogenase. The leuco dye serves as a sink, reacting with the acetaldehyde to form a Schiff base which is intensely red. Thus, the plates can be streaked with a strain or, a strain can be applied in patches to the plate. Colonies will appear a deeper intensity of red dependent upon the level of ADH present in the cell. In particular, by plating appropriate controls on each plate, it is relatively easy to visually discern a strain which has a high level of dehydrogenase (deep red staining), an intermediate level of dehydrogenase (more moderate red staining), and no activity (little or no red staining).

This method thus provides a plate screen that can be employed in the method of the invention.

EXAMPLE 3: Filter screen for HLADH activity.

This example describes a sensitive plate assay of ADH activity which also allows colonies to be tested under different treatment conditions.

This assay relies for manipulation of bacterial colonies on the binding of the colonies to a nitrocellulose filter. The assay is carried out by a modified protocol described by Rellos et al. (Rellos et al., Protein Expression and Purification, 5, 270-277 (1994)). Namely, a series of temperatures between 65 and 85°C in 5°C increments with incubation times varying from 10 minutes to one hour is analyzed in an attempt to determine the cutoff of the stability of the HLADH protein. For these experiments, the source of the *adh* gene encoding the HLADH enzyme was plasmid pBPP (Park et al., J. Biol. Chem., 266, 13296-13302 (1991)).

E. coli DH5α cells containing plasmid pBPP (i.e., HLADH$^+$) or plasmid pCRII (i.e., HLADH$^-$) (InVitrogen; Carlsbad, CA) were grown on rich media plates at cell densities up to about 1,000 colonies per plate and transferred onto a nitrocellulose membrane. The adhered cells were lysed in Buffer 1 (10 mM KMes, pH 6.5, 0.5 mM $CoCl_2$, 0.1% Triton X-100, 50 μg/ml lysozyme, 10 μg/ml DNAse) in a chloroform bath for about one hour, washed once in Buffer 2 (10 mM KMes, 0.5 mM $CoCl_2$, 0.2% BSA), and then washed two more times in Buffer 3 (Buffer 2 without BSA). The filters were then incubated at high temperatures in Buffer 4 (10 mM glycine, 0.5 mM $CoCl_2$) and, after washing in Buffer 3, were incubated in the enzyme-detecting solution (30 mM Tris, pH 8.3, 2% ethanol, 1 mM NAD^+, 0.1 mg/ml phenazine methosulfate, 1 mg/ml nitroblue tetrazolium) at room temperature for 3-5 minutes.

Results of these experiments are depicted in **Figure 2**. As can be seen in this figure, the experiments confirm that a 15-20 minute treatment of the filters at 75°C resulted in roughly 90% inactivation of the HLADH

protein as estimated by the color changes. This information on the activity of the native protein can be used as a baseline for the identification and isolation of mutagenized candidates having altered ADH activity according to the invention.

EXAMPLE 4: Shuttle vectors and use of a *p*-rosaniline assay for verification of the activity of the HLADH gene in *Thermus*

In order to allow expression of the HLADH gene in both *Thermus* and *E. coli*, the gene was subcloned into the *Thermus* shuttle vector, pTG100kan^{tr2} to create plasmid pTG450 depicted in **Figure** 3. In this construct, the gene is placed upstream of the thermostable kanamycin resistance gene (*kan*^{tr2}), which is commanded by the *lac* promoter in *E. coli*, and the *leu* promoter in *Thermus*.

An *E. coli* strain harboring pTG450 has three times more HLADH activity in the presence of IPTG than the strain harboring the original pBPP plasmid. When transformed into *Thermus*, the *adh* gene integrates into the *leu*B site in the *Thermus* chromosome by a double recombination event. For these experiments, *Thermus flavus* was transformed with both the HLADH$^-$ plasmid pTG100kan^{tr2} (i.e., creating strain TGF353) and the HLADH$^+$ plasmid TG450 (i.e., creating strain TGF650).

The presence of the *adh* gene in TGF650 was confirmed by PCR, and both TGF353 and TGF650 cells were assayed using a variation of the *p*-rosaniline plate assay described in Example 2. Namely, the agar overlay contained the same ingredients described, except TT media (Weber et al., Bio/Technology, 13, 271-275 (1995); Oshima et al., International Journal of Systematic Bacteriology, 24, 102-112 (1974)) was employed instead of Luria broth. A standard *p*-rosaniline plate can not be used since the indicator dye will spontaneously convert to the Schiff base if incubated overnight in the plate as part of this assay.

Using this approach, HLADH activity was observed in the pTG450 *Thermus* transformants at a level well above background levels observed for the pTG100kan^{tr2} *Thermus* transformants. The activity was observed up to 70°C. These results thus confirm that a *p*-rosaniline plate assay similarly can be employed in the context of the present invention for screening in *Thermus* for mutants having altered ADH activity.

EXAMPLE 5: Development of a Method of HLADH Selection/Enrichment in *E. coli*

This example describes a method of negative selection for growth of *E. coli* strains harboring the *adh* gene.

For these experiments, *E. coli* DH5α cells containing either pTG100kan^{tr2} (i.e., HLADH$^-$) or pTG450 (i.e., HLADH$^+$) were grown on LB plates with different alcohols in concentrations ranging from 2% to 12%. The results of one such experiment are displayed in **Table 1.**

Table 1. Effect of varying concentrations of alcohol in *Escherichia coli* strain DH5α containing either containing or not containing a plasmid

	% of Methanol					% of Ethanol					% of Isopropanol					% of Propanol			
DH5α	2	4	8	10	12	2	4	8	10	12	2	4	6	8	12	2	4	8	12
PTG100kan^{tr2}	++	++	++	++	++	++	++	++	-	-	+	++	+-	-	-	++	-	-	-
PTG450	++	++	+	+	+-	++	++	+-	-	-	+	+-	-	-	-	+-	-	-	-

Symbols in order of decreasing growth: ++, +, +-, -

As can be seen from **Table 1**, *E. coli* cells harboring high activity of HLADH (i.e., transformed with the HLADH' plasmid pTG450) are more sensitive to the presence of the alcohols in high concentrations. This probably is due to the accumulation of toxic aldehyde levels in the cells which result from the alcohol dehydrogenase reaction. Three other alcohols were tested (i.e., benzyl alcohol, hexyl alcohol, and hexyl amine), but did not give clear results because of their poor solubility in the media.

The experiment was repeated several times and the alcohol levels were refined to determine a range resulting in a clear selection. Three of the alcohols, i.e., ethanol at a concentration of 10%, isopropanol at a concentration of 4%, and propanol at a concentration of 2%, resulted in clean, negative selection for growth of *E. coli* harboring the *adh* gene.

These results thus confirm that the selection scheme can be employed for the isolation of mutants with altered ADH activity and, in particular, to select against *E. coli* strains having high levels of ADH. Such a system of negative selection also can be employed to affirmatively identify mutants having high levels of ADH. For instance, cells can be replica plated onto a series of plates from a single master plate prior to their transfer to nitrocellulose membranes. One of the plates can be retained, instead of being transferred to nitrocellulose, and matched against the sensitive cells identified in the assay. Cells of interest can then be recovered from the untreated plates.

EXAMPLE 6: Development of a Method of HLADH Selection/Enrichment in *Thermus*

This example describes the growth of *Thermus* strains in the presence of the high concentrations of

alcohols as a general method for selecting for growth of *Thermus* strains having high levels of ADH activity.

A series of experiments was conducted to develop a selection using alcohol levels in *Thermus*. In these experiments, *Thermus flavus* strains TGF353 ($HLADH^-$) and TGF670 ($HLADH^+$) were employed. Each strain was grown for two days on *Thermus* rich media (e.g., TT media, as described in Oshima et al., International Journal of Systematic Bacteriology, 24, 102-112 (1974)) present in plates or, was grown overnight in 4 ml of liquid TT medium, in order to ensure the cells were at the same physiological stage prior to testing. The test itself was performed on TT media and *Thermus* minimal media (Yeh et al., J. Biol. Chem., 251, 3134-3139 (1976) containing Casaminoacids (TMIN, CAA). Over a series of many experiments, the strains were grown on agar plates or in liquid medium containing various concentrations of ethanol (i.e., 0.5, 1, 2, 4, 6, or 8%), various concentrations of methanol (i.e., 2, 4, 6, or 8%), various concentrations of isopropanol (i.e., 0.5, 1, 2, 4, or 6%), various concentrations of propanol (i.e., 1, 2, 4, or 6%), or various concentrations of propanediol (i.e. 0.5 or 1%). Such experiments further were done at different pHs, i.e., at pH 7.0, 7.5 and 8.0, for the various alcohols at different concentrations. The results of one of these experiments is set out in **Table 2.**

Table 2. Optical density (OD_{600}) in various media

Strain	pH 7.0				pH 7.0				pH 7.5				pH 7.5			
	% Ethanol				% Isopropanol				% Ethanol				% Isopropanol			
	--	0.5	1	2	0.5	1	2		--	0.5	1	2	0.5	1	2	
TGF353 (ADH')	2.6	1.6	0.1	0.2	0.2	0.1	0.1		2.4	2.2	1.1	0.2	1.9	0.2	0.04	
TGF670 (ADH')	2.2	1.2	1.6	0.3	1.9	0.6	0.1		1.9	1.9	1.6	0.3	1.8	2.2	0.2	

As can be seen from this experiment, the HLADH$^+$ strain TGF670 demonstrates higher resistance to alcohols than the HLADH$^-$ strain TGF353. Moreover, this selection appears to be dependent on pH, with the selection functioning better at lower pH, especially with ethanol. The selection thus may work by lowering the pH of the media--*Thermus* prefers higher pH for growth, in the range of pH 7.5-8.5 -- although not enough *Thermus* biochemistry is known to make this conclusive.

A similar effect can also be achieved on plates. However, the primary effect of the screen in *Thermus* is to retard growth of cells without the *adh* gene, not to completely eliminate it. This also is the case with the liquid media, indicating that a completely clean selection in *Thermus* without background is difficult to achieve. Nevertheless, this selection means provides a powerful enrichment, especially in liquid, by selecting for faster growing cells under the conditions defined.

The results thus confirm that the enrichment/selection means outlined above can be employed with *Thermus*.

EXAMPLE 7: Hydroxylamine mutagenesis of the *adh* gene.

This example describes mutagenesis of the *adh* gene as a representative alcohol dehydrogenase gene using the mutagen hydroxylamine (HA).

For HA mutagenesis of the *adh* gene, plasmids pBPP and pTG450, both of which contain this gene, were treated with HA using a standard approach. Namely, approximately 8 μg of plasmid DNA was mixed with 0.5 M NH_2OH and incubated at 37°C for various lengths of time. For example, aliquots were taken at 1, 2, 3, or 4 hours following treatment, or following overnight exposure to the mutagen. The plasmid DNA was then transformed into *E. coli* strain DH5α and plated onto LB_{Ap100} plates (i.e. LB plates containing 100 μg/ml ampicillin). Transformants

were analyzed by the ADH filter assay described in Example 3, and also using the *p*-rosaniline assay described in Example 2 to estimate the efficiency of mutagenesis.

After overnight treatment, only 3 - 4% plasmids treated with HA remained active. Plasmids treated by HA under conditions providing ~50% of inactivation of the *adh* were then transformed into *E. coli* strain NM554 (obtained from New England Biolabs) to obtain 500 - 700 transformant colonies per plate. These colonies were analyzed by the nitrocellulose filter ADH assay described in Example 3. For heat inactivation of ADH, the filters were incubated for 15 minutes at 70 ^{o}C in a hybridization oven.

Approximately 20,000 transformants were screened using this rapid method. Eighteen candidates were identified which appeared to show increased ADH thermotolerance. The candidates were purified and assayed on the same filter as control strains (i.e., strain XL1 containing the LADH^{+} plasmid pBPP, and strain NM554 containing the LADH^{-} plasmid pBluescript).

Based on results of the filter screening, none of the identified candidates appeared to have the temperature-resistant phenotype suggested by the results of the ADH filter assay. It is possible, however, that thermoresistant mutants can be obtained with HA upon further screening. Moreover, the chances of obtaining mutagenized *adh* resulting in enzyme thermostabilization might be further increased by excising the mutagenized gene from the vector, and resubcloning into a wild-type vector (i.e., a vector that has not been treated with HA), followed by screening.

EXAMPLE 8: PCR Mutagenesis of the *adh* gene

This example describes PCR mutagenesis of the *adh* gene as a representative alcohol dehydrogenase gene.

To increase the efficiency of the cloning of mutagenized *adh*, primers for directional cloning were employed:

CCC CGA ATT CTC AAA ACG TCA GGA TGG TAC G ADH(*Eco*RI) [SEQ ID NO:21]

CCC CTC TAG AAT AAA TGA GCA CAG CAG GAA AAG TAA TAA AAT GC

ADH(*Xba*I) [SEQ ID NO:22]

The *adh* gene was amplified using these primers and cloned into a pGEM-T vector.

For PCR mutagenesis two protocols were used, one according to Spee et al. (Spee et al., Nucl. Acids Res., 21, 777-778 (1993)), and another according to Rellos et al., (Rellos et al., supra) in which the limiting dNTP concentration was double that of the first procedure and dITP was not employed. The pGEM-T plasmid containing the *adh* gene was then used as a template for PCR mutagenesis of *adh* using standard T7 and SP6 primers to perform the error-prone PCR reaction under these conditions.

Mutagenized *adh*-containing fragments were digested using *Xba*I and *Eco*RI enzymes, and subcloned into pBluescript SK to create a pBlue-ADH library. The resultant pBlue-ADH library (i.e., one library for each mutagenesis method performed) was transformed *en masse* into *E. coli* strain NM554 to allow the *adh* gene to be transcribed from the *lac* promoter. Transformants were then analyzed: (i) by PCR to determine the efficiency of cloning (% of the plasmids with and without insert), and ii) by ADH filter assay to determine the efficiency of mutagenesis (% inactive ADH clones). The results of these analyses are shown in **Table 3**.

Table 3. Mutant candidates identified

Method of mutagenesis*	Percentage of the plasmids with the insert	Percentage of the ADH$^+$ clones
Method No. 1	60%	64%
Method No. 2	90%	36%
No mutagenesis (wild-type *adh*)	80%	75%

* Method No.1 was done according to Spee et al., supra, (i.e. with 14 μM of limiting dNTP and 200 μM dITP) and Method No. 2 was done according to Rellos et al., supra (i.e. without dITP and with 25 μM of the limiting dNTP).

As can be seen from these results, both the cloning and mutagenesis efficiency was better using the second method.

The transformants were then plated to a density of 500 - 700 cells per plate and assayed on the filters under the same conditions described in the prior example for HA-mutagenesis of the *adh* gene. Approximately 5,000 clones containing *adh* mutagenized by the first method, and the same number of clones mutagenized by the second method, were tested. No thermostable candidates from the first method were identified. By contrast, thirteen candidates were selected from clones mutagenized by the second method which appeared to possess an HLADH variant that was more stable than the wild-type enzyme. Upon restreaking and retesting these colonies by the filter assay method, nine of the thirteen candidates (i.e., plasmids pAD7, pAD8, pAD10, pAD91, pAD92, pAD93, pAD95, pAD111, and pAD113) were chosen for further characterization.

These results confirm that PCR-mediated mutagenesis, particularly as described herein, can be employed to

obtain potential thermostable LADH variants. The results further indicate that the method can be employed to obtain other stabilized alcohol dehydrogenases, or other stabilized proteins.

EXAMPLE 9: Characterization of thermotolerant HLADH candidates.

This example describes a characterization for increased thermostability of mutants identified in the prior example.

These experiments were done by calculating the residual HLADH activity at 70°C for a series of incubation periods. Residual activity is calculated as activity after incubation at a particular temperature divided by activity before incubation. Cultures of the mutant candidates as well as control cells harboring the wild-type HLADH$^+$ control plasmid pBPP and HLADH$^-$ negative control plasmid pGEM-T were grown in appropriate media, and cell extracts were made by sonication. The extracts were then incubated at 70°C, taking an initial sample (t_0), and sampling at about 30, 60, and 120 minutes. The samples were stored on ice, and the HLADH activity was determined spectrophotometrically as described in Example 1. The data was plotted as a percentage of activity compared to the t_0 activity (residual activity) in order to compare the individual samples to each other and adjust for variations in expression levels or growth variations.

Figure 4 displays the residual activity data for the nine candidate plasmids pAD7, pAD8, pAD10, pAD91, pAD92, pAD93, pAD95, pAD111, and pAD113, wherein the t_0 activity is normalized to 1.00 (100%). As can be seen from **Figure 4**, all the mutants exhibited increased thermotolerance compared to cells containing plasmid pBPP, which contains the wild-type HLADH gene. In particular, plasmids pAD91, pAD92, and pAD10 showed the most noticeable alterations

in thermostability. Cells containing pGEM-T (i.e., not having an HLADH gene) did not show any HLADH activity.

These results thus confirm that the method of the invention can be employed to obtain thermostable alcohol dehydrase, particularly HLADH, mutants.

Table 4 below provides data illustrating comparative data for HALDH activities in the original wild-type ("WT") clone and mutants. All clones were grown in 50 ml of LB medium with 100 μg/ml Amp (12.5 μg/ml Tet for WT clone) overnight, concentrated in 1 ml of the assay buffer (83 mM KH_2PO_4, 40 mM KCl, 0.25 mM EDTA), sonicated and assayed with ethanol as a substrate and NAD cofactor, with results shown as U = mol/mg protein x 1000 / percent residual activity.

Table 4. HALDH Activity after Heat Treatment

	Heat Treatment time			
Strain	RT	15 min	30 min	60 min
pADH7	8/100%	4/50%	2/25%	0.6/8%
pADH8	21/100%	7.4/35%	2/10%	0.2/1%
pADH10	16/100%	4/25%	1.4/9%	0/0%
pADH91	11/100%	8/73%	6/55%	4/36%
pADH92	25/100%	15/60%	17/68%	12/48%
pADH93	6/100%	1/17%	2.5/42%	0/0%
pADH95	66/100%	21/32%	10/15%	3/5%
pADH111	22/100%	15/68%	16/73%	11/50%
pADH113	9/100%	4/44%	3/33%	0.8/9%
WT	10/100%	1/10%	0.3/3%	0/0%

Table 5 below provides data illustrating comparative data for HALDH activities of the original wild-type ("WT") clone and mutants and substrate specificity. All clones were grown in 1 L of LB medium with 100 μg/ml Amp (12.5 μg/ml Tet for WT clone) overnight, concentrated in 50 ml of the assay buffer (83 mM KH_2PO_4, 40 mM KCl, 0.25 mM EDTA), sonicated, incubated at 55°C for 5 min to denature the *E.coli* protiens and lyophilized. The assays were performed at room temperature with the listed substrate and NAD cofactor, with results shown as U = mol/mg protein x 1000.

Table 5 HLADH Substrate Specificity

Strain	Ethanol	Isopropanol	Butanol	Benzyl Alcol
pADH7	8.7	0	5.3	1.4
pADH8	18.2	1.4	11	7
pADH10	15.6	3	11.5	4.7
pADH91	13.2	1.1	4.7	3.4
pADH92	23.5	2.3	11	6.8
pADH93	5.6	1	3	1.6
pADH95	48	0.7	21.3	4.5
pADH111	22.6	1.6	9.8	3
pADH113	7	1.1	3	5
WT	9.2	1.7	7.6	3.5

Strain	Hexanol	Cyclohexanol	R-(-)Butanol	S-(+)Butanol
pADH7	4	3	0	0
pADH8	15	49	2.4	2.2
pADH10	15	69	10	4
pADH91	5.8	23	2.4	1.7
pADH92	10.6	50	2.3	2.4
pADH93	3.9	22	2	1.4
pADH95	21.3	16.5	0.5	0.8
pADH111	9.4	58	4	2.7
pADH113	2.7	14.7	2	1.3
WT	10	42	4.3	2.9

EXAMPLE 10: Sequence Analysis of HLADH Thermotolerant Candidates

This examples describes the sequencing of the mutagenized *adh* genes.

The inserts of plasmids containing the mutagenized *adh* gene were sequenced using an ABI DNA sequencer, and compared to the sequence of the wild type protein. The translated nucleic acid/amino acid sequence for plasmids having the wild-type or mutant *adh* genes is given in **Figure 5**, with the positions of the non-silent mutations (i.e., those that change the encoded amino acid) indicated by the boxes. **Table 6** summarizes all the nucleic acid mutations and the respective amino acid changes, if any, introduced by the mutations.

Table 6. Mutations identified in thermotolerant candidates

Mutant plasmid	Base pair position	Amino acid position[1]	Original codon	Mutant codon	Amino acid change[2]
pAD7	774	257	ATG	ATA	Met257Ile
	878	292	GTG	GCG	Val292Ala
pAD8	285	94	ACT	ACC	no aa change
	806	268	GTC	GCC	Val268Ala
pAD10	227	75	AGC	AAC	Ser75Asn
pAD91/92	284	94	ACT	ATT	Thr94Ile
pAD93	847	282	TGT	AGT	Cys282Ser
	893	297	GAT	GGT	Asp297Gly
pAD95	774	257	ATG	ATA	Met257Ile
	878	292	GTG	GCG	Val292Ala
pAD111	532	177	TCT	ACT	Ser177Thr
pAD113	129	42	GCC	GCT	no aa change
	159	52	GTG	GTA	no aa change
	331	110	TTC	CTC	Phe110Leu

Also, the individual sequences of the mutant *adh* sequences are set forth in the Sequence Listing for pAD7 (i.e., nucleic acid sequence at SEQ ID NO:3 and amino acid sequence at SEQ ID NO:4), pAD8 (i.e., nucleic acid sequence at SEQ ID NO:5 and amino acid sequence at SEQ ID NO:6), pAD10 (i.e., nucleic acid sequence at SEQ ID NO:7 and amino acid sequence at SEQ ID NO:8), pAD91/pAD92 (i.e., nucleic acid sequence at SEQ ID NO:9 and amino acid sequence at SEQ ID NO:10), pAD93 (i.e., nucleic acid sequence at SEQ ID NO:11 and amino acid sequence at SEQ ID NO:12), pAD95 (i.e., nucleic acid sequence at SEQ ID NO:13 and amino acid sequence at SEQ ID NO:14), pAD111

(i.e., nucleic acid sequence at SEQ ID NO:15 and amino acid sequence at SEQ ID NO:16), and pAD113(i.e., nucleic acid sequence at SEQ ID NO:17 and amino acid sequence at SEQ ID NO:18).

The first numbered amino acid in the wild-type and mutant sequences is serine since, in the sequences studied, the initial methionine (Met) is not present in the final protein. However, it is possible that Met is present in the wild-type (or mutant) HLADH sequences that are produced in a different host, e.g., in a eukaryotic host, or when transcribed and translated from a different plasmid construct or chromosome.

As can be seen from this data, the sequences of pAD91 and pAD92 are identical, which indicates the clones from which the DNA was isolated likely are siblings. Mutants containing plasmids pAD91, PAD92, pAD93, and pAD95 were identified from the same filter and mutants containing plasmids pAD111 and pAD113 were identified from the same filter assay. Also, in both pAD8 and pAD91/92, the coding sequence specifying amino acid 94 is mutated. Whereas this results in no change in this position in pAD8, a mutation is introduced here in pAD91/92. Similarly, two mutations in pAD113 are silent and do not produce an amino acid change. These silent mutations likely do not contribute substantially to the thermostability of the protein.

EXAMPLE 11: Further thermostabilization of HLADH proteins

This example describes the means by which the thermostable proteins identified and characterized as in the prior examples can be further thermostabilized.

Using the new mutants as a starting point, the process applied here can be reiterated to increase the thermostability of the HLADH enzyme even further. Namely, it is expected that combinations of the identified HLADH mutations or, combinations of these

mutations with other HLADH mutations, can further thermostabilize the enzyme.

In order to do this, the new thermoinactivation limits need to be defined as described in Example 3. This is followed by a new round of mutagenesis performed as described in Examples 8, 9, and 10. In addition, the identified mutations can be put together in differing combinations by *in vitro* site-directed mutagenesis and further molecular biology methods (see, e.g., Sambrook et al., Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory Press, NY. 1989)) that include DNA shuffling via PCR methods (Stemmer et al., Proc. Natl. Acad. Sci., 91, 10747-10751 (1994a); Stemmer et al., Nature, 340, 389-391 (1994b)). As they have done in the past, these methods are all expected to give further increases in the levels of thermostability of the enzyme or, in another similarly screened-for trait.

All of the references cited herein, including patents, patent applications, sequences, and publications, are hereby incorporated in their entireties by reference.

While this invention has been described with an emphasis upon preferred embodiments, it will be obvious to those of ordinary skill in the art that variations in the preferred embodiments can be used, including variations due to improvements in the art, and that the invention can be practiced otherwise than as specifically described herein. Accordingly, this invention includes all modifications encompassed within the spirit and scope of the invention as defined by the following claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: DAVID C. DEMIRJIAN
IGOR A. BRIKUN
MALCOLM J. CASADABAN
VERONIKA VONSTEIN

(ii) TITLE OF INVENTION: Method For The Stabilization Of Proteins And The Thermostabilized Alcohol Dehydrogenases Produced Thereby

(iii) NUMBER OF SEQUENCES: 24

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Mcdonald Boehnen Hulbert & Berghoff
(B) STREET: 300 South Wacker Drive
(C) CITY: Chicago
(D) STATE: Illinois
(E) COUNTRY: United States
(F) ZIP: 60606

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1128 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
ATG AGC ACA GCA GGA AAA GTA ATA AAA TGC AAA GCG GCT GTG CTG TGG      48
    Ser Thr Ala Gly Lys Val Ile Lys Cys Lys Ala Ala Val Leu Trp
      1               5                  10                  15

GAG GAA AAG AAA CCA TTT TCC ATC GAG GAG GTG GAG GTT GCA CCC CCG      96
Glu Glu Lys Lys Pro Phe Ser Ile Glu Glu Val Glu Val Ala Pro Pro
                20                  25                  30

AAG GCC CAT GAA GTC CGT ATA AAG ATG GTG GCC ACA GGA ATT TGT CGC     144
Lys Ala His Glu Val Arg Ile Lys Met Val Ala Thr Gly Ile Cys Arg
            35                  40                  45

TCA GAT GAC CAC GTG GTT AGT GGA ACC CTT GTC ACA CCT CTT CCT GTG     192
Ser Asp Asp His Val Val Ser Gly Thr Leu Val Thr Pro Leu Pro Val
        50                  55                  60

ATC GCA GGC CAT GAG GCA GCG GGC ATT GTG GAG AGC ATT GGA GAA GGC     240
Ile Ala Gly His Glu Ala Ala Gly Ile Val Glu Ser Ile Gly Glu Gly
    65                  70                  75

GTC ACT ACA GTA AGA CCA GGT GAT AAA GTC ATC CCA CTC TTT ACT CCC     288
Val Thr Thr Val Arg Pro Gly Asp Lys Val Ile Pro Leu Phe Thr Pro
80                  85                  90                  95

CAG TGT GGA AAA TGC AGG GTT TGT AAG CAC CCT GAA GGC AAC TTC TGC     336
Gln Cys Gly Lys Cys Arg Val Cys Lys His Pro Glu Gly Asn Phe Cys
                100                 105                 110

TTG AAA AAT GAT CTG AGC ATG CCT CGG GGA ACC ATG CAG GAT GGT ACC     384
Leu Lys Asn Asp Leu Ser Met Pro Arg Gly Thr Met Gln Asp Gly Thr
            115                 120                 125

AGC AGG TTC ACC TGC AGA GGG AAG CCC ATC CAC CAC TTC CTT GGC ACC     432
Ser Arg Phe Thr Cys Arg Gly Lys Pro Ile His His Phe Leu Gly Thr
        130                 135                 140
```

```
AGC ACC TTC TCC CAG TAC ACC GTG GTG GAC GAG ATC TCA GTG GCC AAG      480
Ser Thr Phe Ser Gln Tyr Thr Val Val Asp Glu Ile Ser Val Ala Lys
145                 150                 155

ATC GAT GCG GCC TCA CCG CTG GAG AAA GTC TGT CTC ATT GGC TGT GGA      528
Ile Asp Ala Ala Ser Pro Leu Glu Lys Val Cys Leu Ile Gly Cys Gly
160                 165                 170                 175

TTT TCT ACT GGT TAT GGG TCT GCA GTC AAG GTT GCC AAG GTC ACC CAG      576
Phe Ser Thr Gly Tyr Gly Ser Ala Val Lys Val Ala Lys Val Thr Gln
                180                 185                 190

GGC TCC ACC TGT GCC GTG TTT GGC CTT GGA GGA GTG GGC CTG TCT GTT      624
Gly Ser Thr Cys Ala Val Phe Gly Leu Gly Gly Val Gly Leu Ser Val
            195                 200                 205

ATC ATG GGC TGT AAA GCA GCC GGA GCG GCC AGG ATC ATT GGG GTG GAC      672
Ile Met Gly Cys Lys Ala Ala Gly Ala Ala Arg Ile Ile Gly Val Asp
        210                 215                 220

ATC AAC AAA GAC AAG TTT GCA AAG GCC AAA GAA GTG GGT GCC ACT GAG      720
Ile Asn Lys Asp Lys Phe Ala Lys Ala Lys Glu Val Gly Ala Thr Glu
225                 230                 235

TGT GTC AAC CCT CAG GAC TAC AAG AAA CCC ATC CAG GAG GTG CTG ACA      768
Cys Val Asn Pro Gln Asp Tyr Lys Lys Pro Ile Gln Glu Val Leu Thr
240                 245                 250                 255

GAA ATG AGC AAT GGA GGT GTG GAT TTT TCC TTT GAA GTC ATT GGT CGG      816
Glu Met Ser Asn Gly Gly Val Asp Phe Ser Phe Glu Val Ile Gly Arg
                260                 265                 270

CTC GAC ACT ATG GTG ACT GCC TTG TCA TGC TGT CAA GAA GCA TAT GGT      864
Leu Asp Thr Met Val Thr Ala Leu Ser Cys Cys Gln Glu Ala Tyr Gly
            275                 280                 285

GTG AGC GTC ATT GTG GGA GTA CCT CCT GAT TCC CAA AAT CTC TCT ATG      912
Val Ser Val Ile Val Gly Val Pro Pro Asp Ser Gln Asn Leu Ser Met
        290                 295                 300

AAT CCT ATG TTG CTA CTG AGT GGA CGT ACC TGG AAA GGA GCT ATT TTT      960
Asn Pro Met Leu Leu Leu Ser Gly Arg Thr Trp Lys Gly Ala Ile Phe
305                 310                 315

GGC GGT TTT AAG AGT AAA GAT TCT GTC CCC AAA CTT GTG GCC GAT TTT     1008
Gly Gly Phe Lys Ser Lys Asp Ser Val Pro Lys Leu Val Ala Asp Phe
320                 325                 330                 335

ATG GCT AAA AAG TTT GCA CTG GAT CCT TTA ATC ACC CAT GTT TTA CCT     1056
Met Ala Lys Lys Phe Ala Leu Asp Pro Leu Ile Thr His Val Leu Pro
                340                 345                 350

TTT GAA AAA ATA AAT GAA GGA TTT GAC CTG CTT CGC TCT GGA GAG AGT     1104
Phe Glu Lys Ile Asn Glu Gly Phe Asp Leu Leu Arg Ser Gly Glu Ser
            355                 360                 365

ATC CGT ACC ATC CTG ACG TTT TGA                                    1128
Ile Arg Thr Ile Leu Thr Phe
        370
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 374 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
Ser Thr Ala Gly Lys Val Ile Lys Cys Lys Ala Ala Val Leu Trp Glu
 1               5                  10                  15

Glu Lys Lys Pro Phe Ser Ile Glu Glu Val Glu Val Ala Pro Pro Lys
            20                  25                  30

Ala His Glu Val Arg Ile Lys Met Val Ala Thr Gly Ile Cys Arg Ser
        35                  40                  45

Asp Asp His Val Val Ser Gly Thr Leu Val Thr Pro Leu Pro Val Ile
    50                  55                  60
```

```
Ala Gly His Glu Ala Ala Gly Ile Val Glu Ser Ile Gly Glu Gly Val
 65                  70                  75                  80

Thr Thr Val Arg Pro Gly Asp Lys Val Ile Pro Leu Phe Thr Pro Gln
                 85                  90                  95

Cys Gly Lys Cys Arg Val Cys Lys His Pro Glu Gly Asn Phe Cys Leu
            100                 105                 110

Lys Asn Asp Leu Ser Met Pro Arg Gly Thr Met Gln Asp Gly Thr Ser
        115                 120                 125

Arg Phe Thr Cys Arg Gly Lys Pro Ile His His Phe Leu Gly Thr Ser
    130                 135                 140

Thr Phe Ser Gln Tyr Thr Val Val Asp Glu Ile Ser Val Ala Lys Ile
145                 150                 155                 160

Asp Ala Ala Ser Pro Leu Glu Lys Val Cys Leu Ile Gly Cys Gly Phe
                165                 170                 175

Ser Thr Gly Tyr Gly Ser Ala Val Lys Val Ala Lys Val Thr Gln Gly
            180                 185                 190

Ser Thr Cys Ala Val Phe Gly Leu Gly Gly Val Gly Leu Ser Val Ile
        195                 200                 205

Met Gly Cys Lys Ala Ala Gly Ala Ala Arg Ile Ile Gly Val Asp Ile
    210                 215                 220

Asn Lys Asp Lys Phe Ala Lys Ala Lys Glu Val Gly Ala Thr Glu Cys
225                 230                 235                 240

Val Asn Pro Gln Asp Tyr Lys Lys Pro Ile Gln Glu Val Leu Thr Glu
                245                 250                 255

Met Ser Asn Gly Gly Val Asp Phe Ser Phe Glu Val Ile Gly Arg Leu
            260                 265                 270

Asp Thr Met Val Thr Ala Leu Ser Cys Cys Gln Glu Ala Tyr Gly Val
        275                 280                 285

Ser Val Ile Val Gly Val Pro Pro Asp Ser Gln Asn Leu Ser Met Asn
    290                 295                 300

Pro Met Leu Leu Leu Ser Gly Arg Thr Trp Lys Gly Ala Ile Phe Gly
305                 310                 315                 320

Gly Phe Lys Ser Lys Asp Ser Val Pro Lys Leu Val Ala Asp Phe Met
                325                 330                 335

Ala Lys Lys Phe Ala Leu Asp Pro Leu Ile Thr His Val Leu Pro Phe
            340                 345                 350

Glu Lys Ile Asn Glu Gly Phe Asp Leu Leu Arg Ser Gly Glu Ser Ile
        355                 360                 365

Arg Thr Ile Leu Thr Phe
    370
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1128 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
ATG AGC ACA GCA GGA AAA GTA ATA AAA TGC AAA GCG GCT GTG CTG TGG        48
    Ser Thr Ala Gly Lys Val Ile Lys Cys Lys Ala Ala Val Leu Trp
      1               5                  10                  15

GAG GAA AAG AAA CCA TTT TCC ATC GAG GAG GTG GAG GTT GCA CCC CCG        96
Glu Glu Lys Lys Pro Phe Ser Ile Glu Glu Val Glu Val Ala Pro Pro
                20                  25                  30

AAG GCC CAT GAA GTC CGT ATA AAG ATG GTG GCC ACA GGA ATT TGT CGC       144
Lys Ala His Glu Val Arg Ile Lys Met Val Ala Thr Gly Ile Cys Arg
```

```
                35                  40                  45

TCA GAT GAC CAC GTG GTT AGT GGA ACC CTT GTC ACA CCT CTT CCT GTG      192
Ser Asp Asp His Val Val Ser Gly Thr Leu Val Thr Pro Leu Pro Val
     50                  55                  60

ATC GCA GGC CAT GAG GCA GCG GGC ATT GTG GAG AGC ATT GGA GAA GGC      240
Ile Ala Gly His Glu Ala Ala Gly Ile Val Glu Ser Ile Gly Glu Gly
 65                  70                  75

GTC ACT ACA GTA AGA CCA GGT GAT AAA GTC ATC CCA CTC TTT ACT CCC      288
Val Thr Thr Val Arg Pro Gly Asp Lys Val Ile Pro Leu Phe Thr Pro
 80                  85                  90                  95

CAG TGT GGA AAA TGC AGG GTT TGT AAG CAC CCT GAA GGC AAC TTC TGC      336
Gln Cys Gly Lys Cys Arg Val Cys Lys His Pro Glu Gly Asn Phe Cys
                100                 105                 110

TTG AAA AAT GAT CTG AGC ATG CCT CGG GGA ACC ATG CAG GAT GGT ACC      384
Leu Lys Asn Asp Leu Ser Met Pro Arg Gly Thr Met Gln Asp Gly Thr
            115                 120                 125

AGC AGG TTC ACC TGC AGA GGG AAG CCC ATC CAC CAC TTC CTT GGC ACC      432
Ser Arg Phe Thr Cys Arg Gly Lys Pro Ile His His Phe Leu Gly Thr
        130                 135                 140

AGC ACC TTC TCC CAG TAC ACC GTG GTG GAC GAG ATC TCA GTG GCC AAG      480
Ser Thr Phe Ser Gln Tyr Thr Val Val Asp Glu Ile Ser Val Ala Lys
    145                 150                 155

ATC GAT GCG GCC TCA CCG CTG GAG AAA GTC TGT CTC ATT GGC TGT GGA      528
Ile Asp Ala Ala Ser Pro Leu Glu Lys Val Cys Leu Ile Gly Cys Gly
160                 165                 170                 175

TTT TCT ACT GGT TAT GGG TCT GCA GTC AAG GTT GCC AAG GTC ACC CAG      576
Phe Ser Thr Gly Tyr Gly Ser Ala Val Lys Val Ala Lys Val Thr Gln
                180                 185                 190

GGC TCC ACC TGT GCC GTG TTT GGC CTT GGA GGA GTG GGC CTG TCT GTT      624
Gly Ser Thr Cys Ala Val Phe Gly Leu Gly Gly Val Gly Leu Ser Val
            195                 200                 205

ATC ATG GGC TGT AAA GCA GCC GGA GCG GCC AGG ATC ATT GGG GTG GAC      672
Ile Met Gly Cys Lys Ala Ala Gly Ala Ala Arg Ile Ile Gly Val Asp
        210                 215                 220

ATC AAC AAA GAC AAG TTT GCA AAG GCC AAA GAA GTG GGT GCC ACT GAG      720
Ile Asn Lys Asp Lys Phe Ala Lys Ala Lys Glu Val Gly Ala Thr Glu
    225                 230                 235

TGT GTC AAC CCT CAG GAC TAC AAG AAA CCC ATC CAG GAG GTG CTG ACA      768
Cys Val Asn Pro Gln Asp Tyr Lys Lys Pro Ile Gln Glu Val Leu Thr
240                 245                 250                 255

GAA ATA AGC AAT GGA GGT GTG GAT TTT TCC TTT GAA GTC ATT GGT CGG      816
Glu Ile Ser Asn Gly Gly Val Asp Phe Ser Phe Glu Val Ile Gly Arg
                260                 265                 270

CTC GAC ACT ATG GTG ACT GCC TTG TCA TGC TGT CAA GAA GCA TAT GGT      864
Leu Asp Thr Met Val Thr Ala Leu Ser Cys Cys Gln Glu Ala Tyr Gly
            275                 280                 285

GTG AGC GTC ATT GCG GGA GTA CCT CCT GAT TCC CAA AAT CTC TCT ATG      912
Val Ser Val Ile Ala Gly Val Pro Pro Asp Ser Gln Asn Leu Ser Met
        290                 295                 300

AAT CCT ATG TTG CTA CTG AGT GGA CGT ACC TGG AAA GGA GCT ATT TTT      960
Asn Pro Met Leu Leu Leu Ser Gly Arg Thr Trp Lys Gly Ala Ile Phe
    305                 310                 315

GGC GGT TTT AAG AGT AAA GAT TCT GTC CCC AAA CTT GTG GCC GAT TTT     1008
Gly Gly Phe Lys Ser Lys Asp Ser Val Pro Lys Leu Val Ala Asp Phe
320                 325                 330                 335

ATG GCT AAA AAG TTT GCA CTG GAT CCT TTA ATC ACC CAT GTT TTA CCT     1056
Met Ala Lys Lys Phe Ala Leu Asp Pro Leu Ile Thr His Val Leu Pro
                340                 345                 350

TTT GAA AAA ATA AAT GAA GGA TTT GAC CTG CTT CGC TCT GGA GAG AGT     1104
Phe Glu Lys Ile Asn Glu Gly Phe Asp Leu Leu Arg Ser Gly Glu Ser
            355                 360                 365

ATC CGT ACC ATC CTG ACG TTT TGA                                    1128
Ile Arg Thr Ile Leu Thr Phe
```

370

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 374 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
Ser Thr Ala Gly Lys Val Ile Lys Cys Lys Ala Ala Val Leu Trp Glu
 1               5                  10                  15

Glu Lys Lys Pro Phe Ser Ile Glu Glu Val Glu Val Ala Pro Pro Lys
            20                  25                  30

Ala His Glu Val Arg Ile Lys Met Val Ala Thr Gly Ile Cys Arg Ser
        35                  40                  45

Asp Asp His Val Val Ser Gly Thr Leu Val Thr Pro Leu Pro Val Ile
    50                  55                  60

Ala Gly His Glu Ala Ala Gly Ile Val Glu Ser Ile Gly Glu Gly Val
65                  70                  75                  80

Thr Thr Val Arg Pro Gly Asp Lys Val Ile Pro Leu Phe Thr Pro Gln
                85                  90                  95

Cys Gly Lys Cys Arg Val Cys Lys His Pro Glu Gly Asn Phe Cys Leu
            100                 105                 110

Lys Asn Asp Leu Ser Met Pro Arg Gly Thr Met Gln Asp Gly Thr Ser
        115                 120                 125

Arg Phe Thr Cys Arg Gly Lys Pro Ile His His Phe Leu Gly Thr Ser
    130                 135                 140

Thr Phe Ser Gln Tyr Thr Val Val Asp Glu Ile Ser Val Ala Lys Ile
145                 150                 155                 160

Asp Ala Ala Ser Pro Leu Glu Lys Val Cys Leu Ile Gly Cys Gly Phe
                165                 170                 175

Ser Thr Gly Tyr Gly Ser Ala Val Lys Val Ala Lys Val Thr Gln Gly
            180                 185                 190

Ser Thr Cys Ala Val Phe Gly Leu Gly Gly Val Gly Leu Ser Val Ile
        195                 200                 205

Met Gly Cys Lys Ala Ala Gly Ala Ala Arg Ile Ile Gly Val Asp Ile
    210                 215                 220

Asn Lys Asp Lys Phe Ala Lys Ala Lys Glu Val Gly Ala Thr Glu Cys
225                 230                 235                 240

Val Asn Pro Gln Asp Tyr Lys Lys Pro Ile Gln Glu Val Leu Thr Glu
                245                 250                 255

Ile Ser Asn Gly Gly Val Asp Phe Ser Phe Glu Val Ile Gly Arg Leu
            260                 265                 270

Asp Thr Met Val Thr Ala Leu Ser Cys Cys Gln Glu Ala Tyr Gly Val
        275                 280                 285

Ser Val Ile Ala Gly Val Pro Pro Asp Ser Gln Asn Leu Ser Met Asn
    290                 295                 300

Pro Met Leu Leu Leu Ser Gly Arg Thr Trp Lys Gly Ala Ile Phe Gly
305                 310                 315                 320

Gly Phe Lys Ser Lys Asp Ser Val Pro Lys Leu Val Ala Asp Phe Met
                325                 330                 335

Ala Lys Lys Phe Ala Leu Asp Pro Leu Ile Thr His Val Leu Pro Phe
            340                 345                 350

Glu Lys Ile Asn Glu Gly Phe Asp Leu Leu Arg Ser Gly Glu Ser Ile
        355                 360                 365

Arg Thr Ile Leu Thr Phe
```

370

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1128 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
ATG AGC ACA GCA GGA AAA GTA ATA AAA TGC AAA GCG GCT GTG CTG TGG      48
Ser Thr Ala Gly Lys Val Ile Lys Cys Lys Ala Ala Val Leu Trp
 1               5                  10                  15

GAG GAA AAG AAA CCA TTT TCC ATC GAG GAG GTG GAG GTT GCA CCC CCG      96
Glu Glu Lys Lys Pro Phe Ser Ile Glu Glu Val Glu Val Ala Pro Pro
            20                  25                  30

AAG GCC CAT GAA GTC CGT ATA AAG ATG GTG GCC ACA GGA ATT TGT CGC     144
Lys Ala His Glu Val Arg Ile Lys Met Val Ala Thr Gly Ile Cys Arg
        35                  40                  45

TCA GAT GAC CAC GTG GTT AGT GGA ACC CTT GTC ACA CCT CTT CCT GTG     192
Ser Asp Asp His Val Val Ser Gly Thr Leu Val Thr Pro Leu Pro Val
    50                  55                  60

ATC GCA GGC CAT GAG GCA GCG GGC ATT GTG GAG AGC ATT GGA GAA GGC     240
Ile Ala Gly His Glu Ala Ala Gly Ile Val Glu Ser Ile Gly Glu Gly
65                  70                  75

GTC ACT ACA GTA AGA CCA GGT GAT AAA GTC ATC CCA CTC TTT ACC CCC     288
Val Thr Thr Val Arg Pro Gly Asp Lys Val Ile Pro Leu Phe Thr Pro
 80                  85                  90                  95

CAG TGT GGA AAA TGC AGG GTT TGT AAG CAC CCT GAA GGC AAC TTC TGC     336
Gln Cys Gly Lys Cys Arg Val Cys Lys His Pro Glu Gly Asn Phe Cys
            100                 105                 110

TTG AAA AAT GAT CTG AGC ATG CCT CGG GGA ACC ATG CAG GAT GGT ACC     384
Leu Lys Asn Asp Leu Ser Met Pro Arg Gly Thr Met Gln Asp Gly Thr
        115                 120                 125

AGC AGG TTC ACC TGC AGA GGG AAG CCC ATC CAC CAC TTC CTT GGC ACC     432
Ser Arg Phe Thr Cys Arg Gly Lys Pro Ile His His Phe Leu Gly Thr
    130                 135                 140

AGC ACC TTC TCC CAG TAC ACC GTG GTG GAC GAG ATC TCA GTG GCC AAG     480
Ser Thr Phe Ser Gln Tyr Thr Val Val Asp Glu Ile Ser Val Ala Lys
145                 150                 155

ATC GAT GCG GCC TCA CCG CTG GAG AAA GTC TGT CTC ATT GGC TGT GGA     528
Ile Asp Ala Ala Ser Pro Leu Glu Lys Val Cys Leu Ile Gly Cys Gly
160                 165                 170                 175

TTT TCT ACT GGT TAT GGG TCT GCA GTC AAG GTT GCC AAG GTC ACC CAG     576
Phe Ser Thr Gly Tyr Gly Ser Ala Val Lys Val Ala Lys Val Thr Gln
            180                 185                 190

GGC TCC ACC TGT GCC GTG TTT GGC CTT GGA GGA GTG GGC CTG TCT GTT     624
Gly Ser Thr Cys Ala Val Phe Gly Leu Gly Gly Val Gly Leu Ser Val
        195                 200                 205

ATC ATG GGC TGT AAA GCA GCC GGA GCG GCC AGG ATC ATT GGG GTG GAC     672
Ile Met Gly Cys Lys Ala Ala Gly Ala Ala Arg Ile Ile Gly Val Asp
    210                 215                 220

ATC AAC AAA GAC AAG TTT GCA AAG GCC AAA GAA GTG GGT GCC ACT GAG     720
Ile Asn Lys Asp Lys Phe Ala Lys Ala Lys Glu Val Gly Ala Thr Glu
225                 230                 235

TGT GTC AAC CCT CAG GAC TAC AAG AAA CCC ATC CAG GAG GTG CTG ACA     768
Cys Val Asn Pro Gln Asp Tyr Lys Lys Pro Ile Gln Glu Val Leu Thr
240                 245                 250                 255

GAA ATG AGC AAT GGA GGT GTG GAT TTT TCC TTT GAA GCC ATT GGT CGG     816
Glu Met Ser Asn Gly Gly Val Asp Phe Ser Phe Glu Ala Ile Gly Arg
            260                 265                 270
```

```
CTC GAC ACT ATG GTG ACT GCC TTG TCA TGC TGT CAA GAA GCA TAT GGT      864
Leu Asp Thr Met Val Thr Ala Leu Ser Cys Cys Gln Glu Ala Tyr Gly
            275                 280                 285

GTG AGC GTC ATT GTG GGA GTA CCT CCT GAT TCC CAA AAT CTC TCT ATG      912
Val Ser Val Ile Val Gly Val Pro Pro Asp Ser Gln Asn Leu Ser Met
        290                 295                 300

AAT CCT ATG TTG CTA CTG AGT GGA CGT ACC TGG AAA GGA GCT ATT TTT      960
Asn Pro Met Leu Leu Leu Ser Gly Arg Thr Trp Lys Gly Ala Ile Phe
    305                 310                 315

GGC GGT TTT AAG AGT AAA GAT TCT GTC CCC AAA CTT GTG GCC GAT TTT     1008
Gly Gly Phe Lys Ser Lys Asp Ser Val Pro Lys Leu Val Ala Asp Phe
320                 325                 330                 335

ATG GCT AAA AAG TTT GCA CTG GAT CCT TTA ATC ACC CAT GTT TTA CCT     1056
Met Ala Lys Lys Phe Ala Leu Asp Pro Leu Ile Thr His Val Leu Pro
                340                 345                 350

TTT GAA AAA ATA AAT GAA GGA TTT GAC CTG CTT CGC TCT GGA GAG AGT     1104
Phe Glu Lys Ile Asn Glu Gly Phe Asp Leu Leu Arg Ser Gly Glu Ser
            355                 360                 365

ATC CGT ACC ATC CTG ACG TTT TGA                                    1128
Ile Arg Thr Ile Leu Thr Phe
        370
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 374 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
Ser Thr Ala Gly Lys Val Ile Lys Cys Lys Ala Ala Val Leu Trp Glu
 1               5                  10                  15

Glu Lys Lys Pro Phe Ser Ile Glu Glu Val Glu Val Ala Pro Pro Lys
            20                  25                  30

Ala His Glu Val Arg Ile Lys Met Val Ala Thr Gly Ile Cys Arg Ser
        35                  40                  45

Asp Asp His Val Val Ser Gly Thr Leu Val Thr Pro Leu Pro Val Ile
    50                  55                  60

Ala Gly His Glu Ala Ala Gly Ile Val Glu Ser Ile Gly Glu Gly Val
65                  70                  75                  80

Thr Thr Val Arg Pro Gly Asp Lys Val Ile Pro Leu Phe Thr Pro Gln
                85                  90                  95

Cys Gly Lys Cys Arg Val Cys Lys His Pro Glu Gly Asn Phe Cys Leu
            100                 105                 110

Lys Asn Asp Leu Ser Met Pro Arg Gly Thr Met Gln Asp Gly Thr Ser
        115                 120                 125

Arg Phe Thr Cys Arg Gly Lys Pro Ile His His Phe Leu Gly Thr Ser
    130                 135                 140

Thr Phe Ser Gln Tyr Thr Val Val Asp Glu Ile Ser Val Ala Lys Ile
145                 150                 155                 160

Asp Ala Ala Ser Pro Leu Glu Lys Val Cys Leu Ile Gly Cys Gly Phe
                165                 170                 175

Ser Thr Gly Tyr Gly Ser Ala Val Lys Val Ala Lys Val Thr Gln Gly
            180                 185                 190

Ser Thr Cys Ala Val Phe Gly Leu Gly Gly Val Gly Leu Ser Val Ile
        195                 200                 205

Met Gly Cys Lys Ala Ala Gly Ala Ala Arg Ile Ile Gly Val Asp Ile
    210                 215                 220

Asn Lys Asp Lys Phe Ala Lys Ala Lys Glu Val Gly Ala Thr Glu Cys
225                 230                 235                 240
```

```
Val Asn Pro Gln Asp Tyr Lys Lys Pro Ile Gln Glu Val Leu Thr Glu
                245                 250                 255

Met Ser Asn Gly Gly Val Asp Phe Ser Phe Glu Ala Ile Gly Arg Leu
            260                 265                 270

Asp Thr Met Val Thr Ala Leu Ser Cys Cys Gln Glu Ala Tyr Gly Val
        275                 280                 285

Ser Val Ile Val Gly Val Pro Pro Asp Ser Gln Asn Leu Ser Met Asn
    290                 295                 300

Pro Met Leu Leu Leu Ser Gly Arg Thr Trp Lys Gly Ala Ile Phe Gly
305                 310                 315                 320

Gly Phe Lys Ser Lys Asp Ser Val Pro Lys Leu Val Ala Asp Phe Met
                325                 330                 335

Ala Lys Lys Phe Ala Leu Asp Pro Leu Ile Thr His Val Leu Pro Phe
            340                 345                 350

Glu Lys Ile Asn Glu Gly Phe Asp Leu Leu Arg Ser Gly Glu Ser Ile
        355                 360                 365

Arg Thr Ile Leu Thr Phe
    370
```

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1128 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

```
ATG AGC ACA GCA GGA AAA GTA ATA AAA TGC AAA GCG GCT GTG CTG TGG      48
    Ser Thr Ala Gly Lys Val Ile Lys Cys Lys Ala Ala Val Leu Trp
      1               5                  10                  15

GAG GAA AAG AAA CCA TTT TCC ATC GAG GAG GTG GAG GTT GCA CCC CCG      96
Glu Glu Lys Lys Pro Phe Ser Ile Glu Glu Val Glu Val Ala Pro Pro
                20                  25                  30

AAG GCC CAT GAA GTC CGT ATA AAG ATG GTG GCC ACA GGA ATT TGT CGC     144
Lys Ala His Glu Val Arg Ile Lys Met Val Ala Thr Gly Ile Cys Arg
            35                  40                  45

TCA GAT GAC CAC GTG GTT AGT GGA ACC CTT GTC ACA CCT CTT CCT GTG     192
Ser Asp Asp His Val Val Ser Gly Thr Leu Val Thr Pro Leu Pro Val
        50                  55                  60

ATC GCA GGC CAT GAG GCA GCG GGC ATT GTG GAG AAC ATT GGA GAA GGC     240
Ile Ala Gly His Glu Ala Ala Gly Ile Val Glu Asn Ile Gly Glu Gly
    65                  70                  75

GTC ACT ACA GTA AGA CCA GGT GAT AAA GTC ATC CCA CTC TTT ACT CCC     288
Val Thr Thr Val Arg Pro Gly Asp Lys Val Ile Pro Leu Phe Thr Pro
80                  85                  90                  95

CAG TGT GGA AAA TGC AGG GTT TGT AAG CAC CCT GAA GGC AAC TTC TGC     336
Gln Cys Gly Lys Cys Arg Val Cys Lys His Pro Glu Gly Asn Phe Cys
                100                 105                 110

TTG AAA AAT GAT CTG AGC ATG CCT CGG GGA ACC ATG CAG GAT GGT ACC     384
Leu Lys Asn Asp Leu Ser Met Pro Arg Gly Thr Met Gln Asp Gly Thr
            115                 120                 125

AGC AGG TTC ACC TGC AGA GGG AAG CCC ATC CAC CAC TTC CTT GGC ACC     432
Ser Arg Phe Thr Cys Arg Gly Lys Pro Ile His His Phe Leu Gly Thr
        130                 135                 140

AGC ACC TTC TCC CAG TAC ACC GTG GTG GAC GAG ATC TCA GTG GCC AAG     480
Ser Thr Phe Ser Gln Tyr Thr Val Val Asp Glu Ile Ser Val Ala Lys
    145                 150                 155

ATC GAT GCG GCC TCA CCG CTG GAG AAA GTC TGT CTC ATT GGC TGT GGA     528
Ile Asp Ala Ala Ser Pro Leu Glu Lys Val Cys Leu Ile Gly Cys Gly
```

```
160                 165                 170                 175

TTT TCT ACT GGT TAT GGG TCT GCA GTC AAG GTT GCC AAG GTC ACC CAG      576
Phe Ser Thr Gly Tyr Gly Ser Ala Val Lys Val Ala Lys Val Thr Gln
                180                 185                 190

GGC TCC ACC TGT GCC GTG TTT GGC CTT GGA GCA GTG GGC CTG TCT GTT      624
Gly Ser Thr Cys Ala Val Phe Gly Leu Gly Gly Val Gly Leu Ser Val
            195                 200                 205

ATC ATG GGC TGT AAA GCA GCC GGA GCG GCC AGG ATC ATT GGG GTG GAC      672
Ile Met Gly Cys Lys Ala Ala Gly Ala Ala Arg Ile Ile Gly Val Asp
        210                 215                 220

ATC AAC AAA GAC AAG TTT GCA AAG GCC AAA GAA GTG GGT GCC ACT GAG      720
Ile Asn Lys Asp Lys Phe Ala Lys Ala Lys Glu Val Gly Ala Thr Glu
    225                 230                 235

TGT GTC AAC CCT CAG GAC TAC AAG AAA CCC ATC CAG GAG GTG CTG ACA      768
Cys Val Asn Pro Gln Asp Tyr Lys Lys Pro Ile Gln Glu Val Leu Thr
240                 245                 250                 255

GAA ATG AGC AAT GGA GGT GTG GAT TTT TCC TTT GAA GTC ATT GGT CGG      816
Glu Met Ser Asn Gly Gly Val Asp Phe Ser Phe Glu Val Ile Gly Arg
                260                 265                 270

CTC GAC ACT ATG GTG ACT GCC TTG TCA TGC TGT CAA GAA GCA TAT GGT      864
Leu Asp Thr Met Val Thr Ala Leu Ser Cys Cys Gln Glu Ala Tyr Gly
            275                 280                 285

GTG AGC GTC ATT GTG GGA GTA CCT CCT GAT TCC CAA AAT CTC TCT ATG      912
Val Ser Val Ile Val Gly Val Pro Pro Asp Ser Gln Asn Leu Ser Met
        290                 295                 300

AAT CCT ATG TTG CTA CTG AGT GGA CGT ACC TGG AAA GGA GCT ATT TTT      960
Asn Pro Met Leu Leu Leu Ser Gly Arg Thr Trp Lys Gly Ala Ile Phe
    305                 310                 315

GGC GGT TTT AAG AGT AAA GAT TCT GTC CCC AAA CTT GTG GCC GAT TTT     1008
Gly Gly Phe Lys Ser Lys Asp Ser Val Pro Lys Leu Val Ala Asp Phe
320                 325                 330                 335

ATG GCT AAA AAG TTT GCA CTG GAT CCT TTA ATC ACC CAT GTT TTA CCT     1056
Met Ala Lys Lys Phe Ala Leu Asp Pro Leu Ile Thr His Val Leu Pro
                340                 345                 350

TTT GAA AAA ATA AAT GAA GGA TTT GAC CTG CTT CGC TCT GGA GAG AGT     1104
Phe Glu Lys Ile Asn Glu Gly Phe Asp Leu Leu Arg Ser Gly Glu Ser
            355                 360                 365

ATC CGT ACC ATC CTG ACG TTT TGA                                    1128
Ile Arg Thr Ile Leu Thr Phe
        370
```

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 374 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

```
Ser Thr Ala Gly Lys Val Ile Lys Cys Lys Ala Ala Val Leu Trp Glu
 1               5                  10                  15

Glu Lys Lys Pro Phe Ser Ile Glu Glu Val Glu Val Ala Pro Pro Lys
            20                  25                  30

Ala His Glu Val Arg Ile Lys Met Val Ala Thr Gly Ile Cys Arg Ser
        35                  40                  45

Asp Asp His Val Val Ser Gly Thr Leu Val Thr Pro Leu Pro Val Ile
    50                  55                  60

Ala Gly His Glu Ala Ala Gly Ile Val Glu Asn Ile Gly Glu Gly Val
65                  70                  75                  80

Thr Thr Val Arg Pro Gly Asp Lys Val Ile Pro Leu Phe Thr Pro Gln
                85                  90                  95
```

```
Cys Gly Lys Cys Arg Val Cys Lys His Pro Glu Gly Asn Phe Cys Leu
            100                 105                 110

Lys Asn Asp Leu Ser Met Pro Arg Gly Thr Met Gln Asp Gly Thr Ser
        115                 120                 125

Arg Phe Thr Cys Arg Gly Lys Pro Ile His His Phe Leu Gly Thr Ser
    130                 135                 140

Thr Phe Ser Gln Tyr Thr Val Val Asp Glu Ile Ser Val Ala Lys Ile
145                 150                 155                 160

Asp Ala Ala Ser Pro Leu Glu Lys Val Cys Leu Ile Gly Cys Gly Phe
                165                 170                 175

Ser Thr Gly Tyr Gly Ser Ala Val Lys Val Ala Lys Val Thr Gln Gly
            180                 185                 190

Ser Thr Cys Ala Val Phe Gly Leu Gly Gly Val Gly Leu Ser Val Ile
        195                 200                 205

Met Gly Cys Lys Ala Ala Gly Ala Ala Arg Ile Ile Gly Val Asp Ile
    210                 215                 220

Asn Lys Asp Lys Phe Ala Lys Ala Lys Glu Val Gly Ala Thr Glu Cys
225                 230                 235                 240

Val Asn Pro Gln Asp Tyr Lys Lys Pro Ile Gln Glu Val Leu Thr Glu
                245                 250                 255

Met Ser Asn Gly Gly Val Asp Phe Ser Phe Glu Val Ile Gly Arg Leu
            260                 265                 270

Asp Thr Met Val Thr Ala Leu Ser Cys Cys Gln Glu Ala Tyr Gly Val
        275                 280                 285

Ser Val Ile Val Gly Val Pro Pro Asp Ser Gln Asn Leu Ser Met Asn
    290                 295                 300

Pro Met Leu Leu Leu Ser Gly Arg Thr Trp Lys Gly Ala Ile Phe Gly
305                 310                 315                 320

Gly Phe Lys Ser Lys Asp Ser Val Pro Lys Leu Val Ala Asp Phe Met
                325                 330                 335

Ala Lys Lys Phe Ala Leu Asp Pro Leu Ile Thr His Val Leu Pro Phe
            340                 345                 350

Glu Lys Ile Asn Glu Gly Phe Asp Leu Leu Arg Ser Gly Glu Ser Ile
        355                 360                 365

Arg Thr Ile Leu Thr Phe
    370
```

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1128 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

```
ATG AGC ACA GCA GGA AAA GTA ATA AAA TGC AAA GCG GCT GTG CTG TGG      48
    Ser Thr Ala Gly Lys Val Ile Lys Cys Lys Ala Ala Val Leu Trp
      1               5                  10                  15

GAG GAA AAG AAA CCA TTT TCC ATC GAG GAG GTG GAG GTT GCA CCC CCG      96
Glu Glu Lys Lys Pro Phe Ser Ile Glu Glu Val Glu Val Ala Pro Pro
                20                  25                  30

AAG GCC CAT GAA GTC CGT ATA AAG ATG GTG GCC ACA GGA ATT TGT CGC     144
Lys Ala His Glu Val Arg Ile Lys Met Val Ala Thr Gly Ile Cys Arg
            35                  40                  45

TCA GAT GAC CAC GTG GTT AGT GGA ACC CTT GTC ACA CCT CTT CCT GTG     192
Ser Asp Asp His Val Val Ser Gly Thr Leu Val Thr Pro Leu Pro Val
        50                  55                  60
```

```
ATC GCA GGC CAT GAG GCA GCG GGC ATT GTG GAG AGC ATT GGA GAA GGC      240
Ile Ala Gly His Glu Ala Ala Gly Ile Val Glu Ser Ile Gly Glu Gly
     65                  70                  75

GTC ACT ACA GTA AGA CCA GGT GAT AAA GTC ATC CCA CTC TTT ATT CCC      288
Val Thr Thr Val Arg Pro Gly Asp Lys Val Ile Pro Leu Phe Ile Pro
 80                  85                  90                  95

CAG TGT GGA AAA TGC AGG GTT TGT AAG CAC CCT GAA GGC AAC TTC TGC      336
Gln Cys Gly Lys Cys Arg Val Cys Lys His Pro Glu Gly Asn Phe Cys
                100                 105                 110

TTG AAA AAT GAT CTG AGC ATG CCT CGG GGA ACC ATG CAG GAT GGT ACC      384
Leu Lys Asn Asp Leu Ser Met Pro Arg Gly Thr Met Gln Asp Gly Thr
            115                 120                 125

AGC AGG TTC ACC TGC AGA GGG AAG CCC ATC CAC CAC TTC CTT GGC ACC      432
Ser Arg Phe Thr Cys Arg Gly Lys Pro Ile His His Phe Leu Gly Thr
        130                 135                 140

AGC ACC TTC TCC CAG TAC ACC GTG GTG GAC GAG ATC TCA GTG GCC AAG      480
Ser Thr Phe Ser Gln Tyr Thr Val Val Asp Glu Ile Ser Val Ala Lys
    145                 150                 155

ATC GAT GCG GCC TCA CCG CTG GAG AAA GTC TGT CTC ATT GGC TGT GGA      528
Ile Asp Ala Ala Ser Pro Leu Glu Lys Val Cys Leu Ile Gly Cys Gly
160                 165                 170                 175

TTT TCT ACT GGT TAT GGG TCT GCA GTC AAG GTT GCC AAG GTC ACC CAG      576
Phe Ser Thr Gly Tyr Gly Ser Ala Val Lys Val Ala Lys Val Thr Gln
                180                 185                 190

GGC TCC ACC TGT GCC GTG TTT GGC CTT GGA GGA GTG GGC CTG TCT GTT      624
Gly Ser Thr Cys Ala Val Phe Gly Leu Gly Gly Val Gly Leu Ser Val
            195                 200                 205

ATC ATG GGC TGT AAA GCA GCC GGA GCG GCC AGG ATC ATT GGG GTG GAC      672
Ile Met Gly Cys Lys Ala Ala Gly Ala Ala Arg Ile Ile Gly Val Asp
        210                 215                 220

ATC AAC AAA GAC AAG TTT GCA AAG GCC AAA GAA GTG GGT GCC ACT GAG      720
Ile Asn Lys Asp Lys Phe Ala Lys Ala Lys Glu Val Gly Ala Thr Glu
    225                 230                 235

TGT GTC AAC CCT CAG GAC TAC AAG AAA CCC ATC CAG GAG GTG CTG ACA      768
Cys Val Asn Pro Gln Asp Tyr Lys Lys Pro Ile Gln Glu Val Leu Thr
240                 245                 250                 255

GAA ATG AGC AAT GGA GGT GTG GAT TTT TCC TTT GAA GTC ATT GGT CGG      816
Glu Met Ser Asn Gly Gly Val Asp Phe Ser Phe Glu Val Ile Gly Arg
                260                 265                 270

CTC GAC ACT ATG GTG ACT GCC TTG TCA TGC TGT CAA GAA GCA TAT GGT      864
Leu Asp Thr Met Val Thr Ala Leu Ser Cys Cys Gln Glu Ala Tyr Gly
            275                 280                 285

GTG AGC GTC ATT GTG GGA GTA CCT CCT GAT TCC CAA AAT CTC TCT ATG      912
Val Ser Val Ile Val Gly Val Pro Pro Asp Ser Gln Asn Leu Ser Met
        290                 295                 300

AAT CCT ATG TTG CTA CTG AGT GGA CGT ACC TGG AAA GGA GCT ATT TTT      960
Asn Pro Met Leu Leu Leu Ser Gly Arg Thr Trp Lys Gly Ala Ile Phe
    305                 310                 315

GGC GGT TTT AAG AGT AAA GAT TCT GTC CCC AAA CTT GTG GCC GAT TTT     1008
Gly Gly Phe Lys Ser Lys Asp Ser Val Pro Lys Leu Val Ala Asp Phe
320                 325                 330                 335

ATG GCT AAA AAG TTT GCA CTG GAT CCT TTA ATC ACC CAT GTT TTA CCT     1056
Met Ala Lys Lys Phe Ala Leu Asp Pro Leu Ile Thr His Val Leu Pro
                340                 345                 350

TTT GAA AAA ATA AAT GAA GGA TTT GAC CTG CTT CGC TCT GGA GAG AGT     1104
Phe Glu Lys Ile Asn Glu Gly Phe Asp Leu Leu Arg Ser Gly Glu Ser
            355                 360                 365

ATC CGT ACC ATC CTG ACG TTT TGA                                    1128
Ile Arg Thr Ile Leu Thr Phe
        370
```

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 374 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

```
Ser Thr Ala Gly Lys Val Ile Lys Cys Lys Ala Ala Val Leu Trp Glu
 1               5                  10                  15

Glu Lys Lys Pro Phe Ser Ile Glu Glu Val Glu Val Ala Pro Pro Lys
            20                  25                  30

Ala His Glu Val Arg Ile Lys Met Val Ala Thr Gly Ile Cys Arg Ser
        35                  40                  45

Asp Asp His Val Val Ser Gly Thr Leu Val Thr Pro Leu Pro Val Ile
    50                  55                  60

Ala Gly His Glu Ala Ala Gly Ile Val Glu Ser Ile Gly Glu Gly Val
65                  70                  75                  80

Thr Thr Val Arg Pro Gly Asp Lys Val Ile Pro Leu Phe Ile Pro Gln
                85                  90                  95

Cys Gly Lys Cys Arg Val Cys Lys His Pro Glu Gly Asn Phe Cys Leu
            100                 105                 110

Lys Asn Asp Leu Ser Met Pro Arg Gly Thr Met Gln Asp Gly Thr Ser
        115                 120                 125

Arg Phe Thr Cys Arg Gly Lys Pro Ile His His Phe Leu Gly Thr Ser
    130                 135                 140

Thr Phe Ser Gln Tyr Thr Val Val Asp Glu Ile Ser Val Ala Lys Ile
145                 150                 155                 160

Asp Ala Ala Ser Pro Leu Glu Lys Val Cys Leu Ile Gly Cys Gly Phe
                165                 170                 175

Ser Thr Gly Tyr Gly Ser Ala Val Lys Val Ala Lys Val Thr Gln Gly
            180                 185                 190

Ser Thr Cys Ala Val Phe Gly Leu Gly Gly Val Gly Leu Ser Val Ile
        195                 200                 205

Met Gly Cys Lys Ala Ala Gly Ala Ala Arg Ile Ile Gly Val Asp Ile
    210                 215                 220

Asn Lys Asp Lys Phe Ala Lys Ala Lys Glu Val Gly Ala Thr Glu Cys
225                 230                 235                 240

Val Asn Pro Gln Asp Tyr Lys Lys Pro Ile Gln Glu Val Leu Thr Glu
                245                 250                 255

Met Ser Asn Gly Gly Val Asp Phe Ser Phe Glu Val Ile Gly Arg Leu
            260                 265                 270

Asp Thr Met Val Thr Ala Leu Ser Cys Cys Gln Glu Ala Tyr Gly Val
        275                 280                 285

Ser Val Ile Val Gly Val Pro Pro Asp Ser Gln Asn Leu Ser Met Asn
    290                 295                 300

Pro Met Leu Leu Leu Ser Gly Arg Thr Trp Lys Gly Ala Ile Phe Gly
305                 310                 315                 320

Gly Phe Lys Ser Lys Asp Ser Val Pro Lys Leu Val Ala Asp Phe Met
                325                 330                 335

Ala Lys Lys Phe Ala Leu Asp Pro Leu Ile Thr His Val Leu Pro Phe
            340                 345                 350

Glu Lys Ile Asn Glu Gly Phe Asp Leu Leu Arg Ser Gly Glu Ser Ile
        355                 360                 365

Arg Thr Ile Leu Thr Phe
    370
```

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1128 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```
ATG AGC ACA GCA GGA AAA GTA ATA AAA TGC AAA GCG GCT GTG CTG TGG        48
    Ser Thr Ala Gly Lys Val Ile Lys Cys Lys Ala Ala Val Leu Trp
      1               5                  10                  15

GAG GAA AAG AAA CCA TTT TCC ATC GAG GAG GTG GAG GTT GCA CCC CCG        96
Glu Glu Lys Lys Pro Phe Ser Ile Glu Glu Val Glu Val Ala Pro Pro
                  20                  25                  30

AAG GCC CAT GAA GTC CGT ATA AAG ATG GTG GCC ACA GGA ATT TGT CGC       144
Lys Ala His Glu Val Arg Ile Lys Met Val Ala Thr Gly Ile Cys Arg
              35                  40                  45

TCA GAT GAC CAC GTG GTT AGT GGA ACC CTT GTC ACA CCT CTT CCT GTG       192
Ser Asp Asp His Val Val Ser Gly Thr Leu Val Thr Pro Leu Pro Val
          50                  55                  60

ATC GCA GGC CAT GAG GCA GCG GGC ATT GTG GAG AGC ATT GGA GAA GGC       240
Ile Ala Gly His Glu Ala Ala Gly Ile Val Glu Ser Ile Gly Glu Gly
      65                  70                  75

GTC ACT ACA GTA AGA CCA GGT GAT AAA GTC ATC CCA CTC TTT ACT CCC       288
Val Thr Thr Val Arg Pro Gly Asp Lys Val Ile Pro Leu Phe Thr Pro
 80                  85                  90                  95

CAG TGT GGA AAA TGC AGG GTT TGT AAG CAC CCT GAA GGC AAC TTC TGC       336
Gln Cys Gly Lys Cys Arg Val Cys Lys His Pro Glu Gly Asn Phe Cys
                100                 105                 110

TTG AAA AAT GAT CTG AGC ATG CCT CGG GGA ACC ATG CAG GAT GGT ACC       384
Leu Lys Asn Asp Leu Ser Met Pro Arg Gly Thr Met Gln Asp Gly Thr
            115                 120                 125

AGC AGG TTC ACC TGC AGA GGG AAG CCC ATC CAC CAC TTC CTT GGC ACC       432
Ser Arg Phe Thr Cys Arg Gly Lys Pro Ile His His Phe Leu Gly Thr
        130                 135                 140

AGC ACC TTC TCC CAG TAC ACC GTG GTG GAC GAG ATC TCA GTG GCC AAG       480
Ser Thr Phe Ser Gln Tyr Thr Val Val Asp Glu Ile Ser Val Ala Lys
    145                 150                 155

ATC GAT GCG GCC TCA CCG CTG GAG AAA GTC TGT CTC ATT GGC TGT GGA       528
Ile Asp Ala Ala Ser Pro Leu Glu Lys Val Cys Leu Ile Gly Cys Gly
160                 165                 170                 175

TTT TCT ACT GGT TAT GGG TCT GCA GTC AAG GTT GCC AAG GTC ACC CAG       576
Phe Ser Thr Gly Tyr Gly Ser Ala Val Lys Val Ala Lys Val Thr Gln
                180                 185                 190

GGC TCC ACC TGT GCC GTG TTT GGC CTT GGA GGA GTG GGC CTG TCT GTT       624
Gly Ser Thr Cys Ala Val Phe Gly Leu Gly Gly Val Gly Leu Ser Val
            195                 200                 205

ATC ATG GGC TGT AAA GCA GCC GGA GCG GCC AGG ATC ATT GGG GTG GAC       672
Ile Met Gly Cys Lys Ala Ala Gly Ala Ala Arg Ile Ile Gly Val Asp
        210                 215                 220

ATC AAC AAA GAC AAG TTT GCA AAG GCC AAA GAA GTG GGT GCC ACT GAG       720
Ile Asn Lys Asp Lys Phe Ala Lys Ala Lys Glu Val Gly Ala Thr Glu
    225                 230                 235

TGT GTC AAC CCT CAG GAC TAC AAG AAA CCC ATC CAG GAG GTG CTG ACA       768
Cys Val Asn Pro Gln Asp Tyr Lys Lys Pro Ile Gln Glu Val Leu Thr
240                 245                 250                 255

GAA ATG AGC AAT GGA GGT GTG GAT TTT TCC TTT GAA GTC ATT GGT CGG       816
Glu Met Ser Asn Gly Gly Val Asp Phe Ser Phe Glu Val Ile Gly Arg
                260                 265                 270

CTC GAC ACT ATG GTG ACT GCC TTG TCA TGC AGT CAA GAA GCA TAT GGT       864
Leu Asp Thr Met Val Thr Ala Leu Ser Cys Ser Gln Glu Ala Tyr Gly
            275                 280                 285

GTG AGC GTC ATT GTG GGA GTA CCT CCT GGT TCC CAA AAT CTC TCT ATG       912
Val Ser Val Ile Val Gly Val Pro Pro Gly Ser Gln Asn Leu Ser Met
```

```
          290                 295                 300

AAT CCT ATG TTG CTA CTG AGT GGA CGT ACC TGG AAA GGA GCT ATT TTT      960
Asn Pro Met Leu Leu Leu Ser Gly Arg Thr Trp Lys Gly Ala Ile Phe
    305                 310                 315

GGC GGT TTT AAG AGT AAA GAT TCT GTC CCC AAA CTT GTG GCC GAT TTT     1008
Gly Gly Phe Lys Ser Lys Asp Ser Val Pro Lys Leu Val Ala Asp Phe
320                 325                 330                 335

ATG GCT AAA AAG TTT GCA CTG GAT CCT TTA ATC ACC CAT GTT TTA CCT     1056
Met Ala Lys Lys Phe Ala Leu Asp Pro Leu Ile Thr His Val Leu Pro
                340                 345                 350

TTT GAA AAA ATA AAT GAA GGA TTT GAC CTG CTT CGC TCT GGA GAG AGT     1104
Phe Glu Lys Ile Asn Glu Gly Phe Asp Leu Leu Arg Ser Gly Glu Ser
            355                 360                 365

ATC CGT ACC ATC CTG ACG TTT TGA                                    1128
Ile Arg Thr Ile Leu Thr Phe
        370
```

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 374 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```
Ser Thr Ala Gly Lys Val Ile Lys Cys Lys Ala Ala Val Leu Trp Glu
 1               5                  10                  15

Glu Lys Lys Pro Phe Ser Ile Glu Glu Val Glu Val Ala Pro Pro Lys
            20                  25                  30

Ala His Glu Val Arg Ile Lys Met Val Ala Thr Gly Ile Cys Arg Ser
        35                  40                  45

Asp Asp His Val Val Ser Gly Thr Leu Val Thr Pro Leu Pro Val Ile
    50                  55                  60

Ala Gly His Glu Ala Ala Gly Ile Val Glu Ser Ile Gly Glu Gly Val
65                  70                  75                  80

Thr Thr Val Arg Pro Gly Asp Lys Val Ile Pro Leu Phe Thr Pro Gln
                85                  90                  95

Cys Gly Lys Cys Arg Val Cys Lys His Pro Glu Gly Asn Phe Cys Leu
            100                 105                 110

Lys Asn Asp Leu Ser Met Pro Arg Gly Thr Met Gln Asp Gly Thr Ser
        115                 120                 125

Arg Phe Thr Cys Arg Gly Lys Pro Ile His His Phe Leu Gly Thr Ser
    130                 135                 140

Thr Phe Ser Gln Tyr Thr Val Val Asp Glu Ile Ser Val Ala Lys Ile
145                 150                 155                 160

Asp Ala Ala Ser Pro Leu Glu Lys Val Cys Leu Ile Gly Cys Gly Phe
                165                 170                 175

Ser Thr Gly Tyr Gly Ser Ala Val Lys Val Ala Lys Val Thr Gln Gly
            180                 185                 190

Ser Thr Cys Ala Val Phe Gly Leu Gly Gly Val Gly Leu Ser Val Ile
        195                 200                 205

Met Gly Cys Lys Ala Ala Gly Ala Ala Arg Ile Ile Gly Val Asp Ile
    210                 215                 220

Asn Lys Asp Lys Phe Ala Lys Ala Lys Glu Val Gly Ala Thr Glu Cys
225                 230                 235                 240

Val Asn Pro Gln Asp Tyr Lys Lys Pro Ile Gln Glu Val Leu Thr Glu
                245                 250                 255

Met Ser Asn Gly Gly Val Asp Phe Ser Phe Glu Val Ile Gly Arg Leu
            260                 265                 270
```

```
Asp Thr Met Val Thr Ala Leu Ser Cys Ser Gln Glu Ala Tyr Gly Val
        275                 280                 285

Ser Val Ile Val Gly Val Pro Pro Gly Ser Gln Asn Leu Ser Met Asn
    290                 295                 300

Pro Met Leu Leu Leu Ser Gly Arg Thr Trp Lys Gly Ala Ile Phe Gly
305                 310                 315                 320

Gly Phe Lys Ser Lys Asp Ser Val Pro Lys Leu Val Ala Asp Phe Met
                325                 330                 335

Ala Lys Lys Phe Ala Leu Asp Pro Leu Ile Thr His Val Leu Pro Phe
            340                 345                 350

Glu Lys Ile Asn Glu Gly Phe Asp Leu Leu Arg Ser Gly Glu Ser Ile
        355                 360                 365

Arg Thr Ile Leu Thr Phe
    370
```

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1128 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

```
ATG AGC ACA GCA GGA AAA GTA ATA AAA TGC AAA GCG GCT GTG CTG TGG      48
    Ser Thr Ala Gly Lys Val Ile Lys Cys Lys Ala Ala Val Leu Trp
      1               5                  10                  15

GAG GAA AAG AAA CCA TTT TCC ATC GAG GAG GTG GAG GTT GCA CCC CCG      96
Glu Glu Lys Lys Pro Phe Ser Ile Glu Glu Val Glu Val Ala Pro Pro
                20                  25                  30

AAG GCC CAT GAA GTC CGT ATA AAG ATG GTG GCC ACA GGA ATT TGT CGC     144
Lys Ala His Glu Val Arg Ile Lys Met Val Ala Thr Gly Ile Cys Arg
            35                  40                  45

TCA GAT GAC CAC GTG GTT AGT GGA ACC CTT GTC ACA CCT CTT CCT GTG     192
Ser Asp Asp His Val Val Ser Gly Thr Leu Val Thr Pro Leu Pro Val
        50                  55                  60

ATC GCA GGC CAT GAG GCA GCG GGC ATT GTG GAG AGC ATT GGA GAA GGC     240
Ile Ala Gly His Glu Ala Ala Gly Ile Val Glu Ser Ile Gly Glu Gly
    65                  70                  75

GTC ACT ACA GTA AGA CCA GGT GAT AAA GTC ATC CCA CTC TTT ACT CCC     288
Val Thr Thr Val Arg Pro Gly Asp Lys Val Ile Pro Leu Phe Thr Pro
80                  85                  90                  95

CAG TGT GGA AAA TGC AGG GTT TGT AAG CAC CCT GAA GGC AAC TTC TGC     336
Gln Cys Gly Lys Cys Arg Val Cys Lys His Pro Glu Gly Asn Phe Cys
                100                 105                 110

TTG AAA AAT GAT CTG AGC ATG CCT CGG GGA ACC ATG CAG GAT GGT ACC     384
Leu Lys Asn Asp Leu Ser Met Pro Arg Gly Thr Met Gln Asp Gly Thr
            115                 120                 125

AGC AGG TTC ACC TGC AGA GGG AAG CCC ATC CAC CAC TTC CTT GGC ACC     432
Ser Arg Phe Thr Cys Arg Gly Lys Pro Ile His His Phe Leu Gly Thr
        130                 135                 140

AGC ACC TTC TCC CAG TAC ACC GTG GTG GAC GAG ATC TCA GTG GCC AAG     480
Ser Thr Phe Ser Gln Tyr Thr Val Val Asp Glu Ile Ser Val Ala Lys
    145                 150                 155

ATC GAT GCG GCC TCA CCG CTG GAG AAA GTC TGT CTC ATT GGC TGT GGA     528
Ile Asp Ala Ala Ser Pro Leu Glu Lys Val Cys Leu Ile Gly Cys Gly
160                 165                 170                 175

TTT TCT ACT GGT TAT GGG TCT GCA GTC AAG GTT GCC AAG GTC ACC CAG     576
Phe Ser Thr Gly Tyr Gly Ser Ala Val Lys Val Ala Lys Val Thr Gln
                180                 185                 190
```

```
GGC TCC ACC TGT GCC GTG TTT GGC CTT GGA GGA GTG GGC CTG TCT GTT      624
Gly Ser Thr Cys Ala Val Phe Gly Leu Gly Gly Val Gly Leu Ser Val
            195                 200                 205

ATC ATG GGC TGT AAA GCA GCC GGA GCG GCC AGG ATC ATT GGG GTG GAC      672
Ile Met Gly Cys Lys Ala Ala Gly Ala Ala Arg Ile Ile Gly Val Asp
        210                 215                 220

ATC AAC AAA GAC AAG TTT GCA AAG GCC AAA GAA GTG GGT GCC ACT GAG      720
Ile Asn Lys Asp Lys Phe Ala Lys Ala Lys Glu Val Gly Ala Thr Glu
    225                 230                 235

TGT GTC AAC CCT CAG GAC TAC AAG AAA CCC ATC CAG GAG GTG CTG ACA      768
Cys Val Asn Pro Gln Asp Tyr Lys Lys Pro Ile Gln Glu Val Leu Thr
240                 245                 250                 255

GAA ATA AGC AAT GGA GGT GTG GAT TTT TCC TTT GAA GTC ATT GGT CGG      816
Glu Ile Ser Asn Gly Gly Val Asp Phe Ser Phe Glu Val Ile Gly Arg
                260                 265                 270

CTC GAC ACT ATG GTG ACT GCC TTG TCA TGC TGT CAA GAA GCA TAT GGT      864
Leu Asp Thr Met Val Thr Ala Leu Ser Cys Cys Gln Glu Ala Tyr Gly
            275                 280                 285

GTG AGC GTC ATT GCG GGA GTA CCT CCT GAT TCC CAA AAT CTC TCT ATG      912
Val Ser Val Ile Ala Gly Val Pro Pro Asp Ser Gln Asn Leu Ser Met
        290                 295                 300

AAT CCT ATG TTG CTA CTG AGT GGA CGT ACC TGG AAA GGA GCT ATT TTT      960
Asn Pro Met Leu Leu Leu Ser Gly Arg Thr Trp Lys Gly Ala Ile Phe
    305                 310                 315

GGC GGT TTT AAG AGT AAA GAT TCT GTC CCC AAA CTT GTG GCC GAT TTT     1008
Gly Gly Phe Lys Ser Lys Asp Ser Val Pro Lys Leu Val Ala Asp Phe
320                 325                 330                 335

ATG GCT AAA AAG TTT GCA CTG GAT CCT TTA ATC ACC CAT GTT TTA CCT     1056
Met Ala Lys Lys Phe Ala Leu Asp Pro Leu Ile Thr His Val Leu Pro
                340                 345                 350

TTT GAA AAA ATA AAT GAA GGA TTT GAC CTG CTT CGC TCT GGA GAG AGT     1104
Phe Glu Lys Ile Asn Glu Gly Phe Asp Leu Leu Arg Ser Gly Glu Ser
            355                 360                 365

ATC CGT ACC ATC CTG ACG TTT TGA                                    1128
Ile Arg Thr Ile Leu Thr Phe
        370
```

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 374 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

```
Ser Thr Ala Gly Lys Val Ile Lys Cys Lys Ala Ala Val Leu Trp Glu
  1               5                  10                  15

Glu Lys Lys Pro Phe Ser Ile Glu Glu Val Glu Val Ala Pro Pro Lys
            20                  25                  30

Ala His Glu Val Arg Ile Lys Met Val Ala Thr Gly Ile Cys Arg Ser
        35                  40                  45

Asp Asp His Val Val Ser Gly Thr Leu Val Thr Pro Leu Pro Val Ile
    50                  55                  60

Ala Gly His Glu Ala Ala Gly Ile Val Glu Ser Ile Gly Glu Gly Val
65                  70                  75                  80

Thr Thr Val Arg Pro Gly Asp Lys Val Ile Pro Leu Phe Thr Pro Gln
                85                  90                  95

Cys Gly Lys Cys Arg Val Cys Lys His Pro Glu Gly Asn Phe Cys Leu
            100                 105                 110

Lys Asn Asp Leu Ser Met Pro Arg Gly Thr Met Gln Asp Gly Thr Ser
        115                 120                 125
```

```
Arg Phe Thr Cys Arg Gly Lys Pro Ile His His Phe Leu Gly Thr Ser
130                 135                 140

Thr Phe Ser Gln Tyr Thr Val Val Asp Glu Ile Ser Val Ala Lys Ile
145                 150                 155                 160

Asp Ala Ala Ser Pro Leu Glu Lys Val Cys Leu Ile Gly Cys Gly Phe
                165                 170                 175

Ser Thr Gly Tyr Gly Ser Ala Val Lys Val Ala Lys Val Thr Gln Gly
            180                 185                 190

Ser Thr Cys Ala Val Phe Gly Leu Gly Gly Val Gly Leu Ser Val Ile
        195                 200                 205

Met Gly Cys Lys Ala Ala Gly Ala Ala Arg Ile Ile Gly Val Asp Ile
    210                 215                 220

Asn Lys Asp Lys Phe Ala Lys Ala Lys Glu Val Gly Ala Thr Glu Cys
225                 230                 235                 240

Val Asn Pro Gln Asp Tyr Lys Lys Pro Ile Gln Glu Val Leu Thr Glu
                245                 250                 255

Ile Ser Asn Gly Gly Val Asp Phe Ser Phe Glu Val Ile Gly Arg Leu
            260                 265                 270

Asp Thr Met Val Thr Ala Leu Ser Cys Cys Gln Glu Ala Tyr Gly Val
        275                 280                 285

Ser Val Ile Ala Gly Val Pro Pro Asp Ser Gln Asn Leu Ser Met Asn
    290                 295                 300

Pro Met Leu Leu Leu Ser Gly Arg Thr Trp Lys Gly Ala Ile Phe Gly
305                 310                 315                 320

Gly Phe Lys Ser Lys Asp Ser Val Pro Lys Leu Val Ala Asp Phe Met
                325                 330                 335

Ala Lys Lys Phe Ala Leu Asp Pro Leu Ile Thr His Val Leu Pro Phe
            340                 345                 350

Glu Lys Ile Asn Glu Gly Phe Asp Leu Leu Arg Ser Gly Glu Ser Ile
        355                 360                 365

Arg Thr Ile Leu Thr Phe
    370
```

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1128 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

```
ATG AGC ACA GCA GGA AAA GTA ATA AAA TGC AAA GCG GCT GTG CTG TGG          48
Ser Thr Ala Gly Lys Val Ile Lys Cys Lys Ala Ala Val Leu Trp
 1               5                  10                  15

GAG GAA AAG AAA CCA TTT TCC ATC GAG GAG GTG GAG GTT GCA CCC CCG          96
Glu Glu Lys Lys Pro Phe Ser Ile Glu Glu Val Glu Val Ala Pro Pro
            20                  25                  30

AAG GCC CAT GAA GTC CGT ATA AAG ATG GTG GCC ACA GGA ATT TGT CGC         144
Lys Ala His Glu Val Arg Ile Lys Met Val Ala Thr Gly Ile Cys Arg
        35                  40                  45

TCA GAT GAC CAC GTG GTT AGT GGA ACC CTT GTC ACA CCT CTT CCT GTG         192
Ser Asp Asp His Val Val Ser Gly Thr Leu Val Thr Pro Leu Pro Val
    50                  55                  60

ATC GCA GGC CAT GAG GCA GCG GGC ATT GTG GAG AGC ATT GGA GAA GGC         240
Ile Ala Gly His Glu Ala Ala Gly Ile Val Glu Ser Ile Gly Glu Gly
65                  70                  75

GTC ACT ACA GTA AGA CCA GGT GAT AAA GTC ATC CCA CTC TTT ACT CCC         288
Val Thr Thr Val Arg Pro Gly Asp Lys Val Ile Pro Leu Phe Thr Pro
```

```
                80                  85                  90                  95

CAG TGT GGA AAA TGC AGG GTT TGT AAG CAC CCT GAA GGC AAC TTC TGC                 336
Gln Cys Gly Lys Cys Arg Val Cys Lys His Pro Glu Gly Asn Phe Cys
                100                 105                 110

TTG AAA AAT GAT CTG AGC ATG CCT CGG GGA ACC ATG CAG GAT GGT ACC                 384
Leu Lys Asn Asp Leu Ser Met Pro Arg Gly Thr Met Gln Asp Gly Thr
            115                 120                 125

AGC AGG TTC ACC TGC AGA GGG AAG CCC ATC CAC CAC TTC CTT GGC ACC                 432
Ser Arg Phe Thr Cys Arg Gly Lys Pro Ile His His Phe Leu Gly Thr
        130                 135                 140

AGC ACC TTC TCC CAG TAC ACC GTG GTG GAC GAG ATC TCA GTG GCC AAG                 480
Ser Thr Phe Ser Gln Tyr Thr Val Val Asp Glu Ile Ser Val Ala Lys
    145                 150                 155

ATC GAT GCG GCC TCA CCG CTG GAG AAA GTC TGT CTC ATT GGC TGT GGA                 528
Ile Asp Ala Ala Ser Pro Leu Glu Lys Val Cys Leu Ile Gly Cys Gly
160                 165                 170                 175

TTT ACT ACT GGT TAT GGG TCT GCA GTC AAG GTT GCC AAG GTC ACC CAG                 576
Phe Thr Thr Gly Tyr Gly Ser Ala Val Lys Val Ala Lys Val Thr Gln
                180                 185                 190

GGC TCC ACC TGT GCC GTG TTT GGC CTT GGA GGA GTG GGC CTG TCT GTT                 624
Gly Ser Thr Cys Ala Val Phe Gly Leu Gly Gly Val Gly Leu Ser Val
            195                 200                 205

ATC ATG GGC TGT AAA GCA GCC GGA GCG GCC AGG ATC ATT GGG GTG GAC                 672
Ile Met Gly Cys Lys Ala Ala Gly Ala Ala Arg Ile Ile Gly Val Asp
        210                 215                 220

ATC AAC AAA GAC AAG TTT GCA AAG GCC AAA GAA GTG GGT GCC ACT GAG                 720
Ile Asn Lys Asp Lys Phe Ala Lys Ala Lys Glu Val Gly Ala Thr Glu
    225                 230                 235

TGT GTC AAC CCT CAG GAC TAC AAG AAA CCC ATC CAG GAG GTG CTG ACA                 768
Cys Val Asn Pro Gln Asp Tyr Lys Lys Pro Ile Gln Glu Val Leu Thr
240                 245                 250                 255

GAA ATG AGC AAT GGA GGT GTG GAT TTT TCC TTT GAA GTC ATT GGT CGG                 816
Glu Met Ser Asn Gly Gly Val Asp Phe Ser Phe Glu Val Ile Gly Arg
                260                 265                 270

CTC GAC ACT ATG GTG ACT GCC TTG TCA TGC TGT CAA GAA GCA TAT GGT                 864
Leu Asp Thr Met Val Thr Ala Leu Ser Cys Cys Gln Glu Ala Tyr Gly
            275                 280                 285

GTG AGC GTC ATT GTG GGA GTA CCT CCT GAT TCC CAA AAT CTC TCT ATG                 912
Val Ser Val Ile Val Gly Val Pro Pro Asp Ser Gln Asn Leu Ser Met
        290                 295                 300

AAT CCT ATG TTG CTA CTG AGT GGA CGT ACC TGG AAA GGA GCT ATT TTT                 960
Asn Pro Met Leu Leu Leu Ser Gly Arg Thr Trp Lys Gly Ala Ile Phe
    305                 310                 315

GGC GGT TTT AAG AGT AAA GAT TCT GTC CCC AAA CTT GTG GCC GAT TTT                 1008
Gly Gly Phe Lys Ser Lys Asp Ser Val Pro Lys Leu Val Ala Asp Phe
320                 325                 330                 335

ATG GCT AAA AAG TTT GCA CTG GAT CCT TTA ATC ACC CAT GTT TTA CCT                 1056
Met Ala Lys Lys Phe Ala Leu Asp Pro Leu Ile Thr His Val Leu Pro
                340                 345                 350

TTT GAA AAA ATA AAT GAA GGA TTT GAC CTG CTT CGC TCT GGA GAG AGT                 1104
Phe Glu Lys Ile Asn Glu Gly Phe Asp Leu Leu Arg Ser Gly Glu Ser
            355                 360                 365

ATC CGT ACC ATC CTG ACG TTT TGA                                                1128
Ile Arg Thr Ile Leu Thr Phe
        370
```

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 374 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Ser Thr Ala Gly Lys Val Ile Lys Cys Lys Ala Ala Val Leu Trp Glu
1 5 10 15

Glu Lys Lys Pro Phe Ser Ile Glu Glu Val Glu Val Ala Pro Pro Lys
20 25 30

Ala His Glu Val Arg Ile Lys Met Val Ala Thr Gly Ile Cys Arg Ser
35 40 45

Asp Asp His Val Val Ser Gly Thr Leu Val Thr Pro Leu Pro Val Ile
50 55 60

Ala Gly His Glu Ala Ala Gly Ile Val Glu Ser Ile Gly Glu Gly Val
65 70 75 80

Thr Thr Val Arg Pro Gly Asp Lys Val Ile Pro Leu Phe Thr Pro Gln
85 90 95

Cys Gly Lys Cys Arg Val Cys Lys His Pro Glu Gly Asn Phe Cys Leu
100 105 110

Lys Asn Asp Leu Ser Met Pro Arg Gly Thr Met Gln Asp Gly Thr Ser
115 120 125

Arg Phe Thr Cys Arg Gly Lys Pro Ile His His Phe Leu Gly Thr Ser
130 135 140

Thr Phe Ser Gln Tyr Thr Val Val Asp Glu Ile Ser Val Ala Lys Ile
145 150 155 160

Asp Ala Ala Ser Pro Leu Glu Lys Val Cys Leu Ile Gly Cys Gly Phe
165 170 175

Thr Thr Gly Tyr Gly Ser Ala Val Lys Val Ala Lys Val Thr Gln Gly
180 185 190

Ser Thr Cys Ala Val Phe Gly Leu Gly Gly Val Gly Leu Ser Val Ile
195 200 205

Met Gly Cys Lys Ala Ala Gly Ala Ala Arg Ile Ile Gly Val Asp Ile
210 215 220

Asn Lys Asp Lys Phe Ala Lys Ala Lys Glu Val Gly Ala Thr Glu Cys
225 230 235 240

Val Asn Pro Gln Asp Tyr Lys Lys Pro Ile Gln Glu Val Leu Thr Glu
245 250 255

Met Ser Asn Gly Gly Val Asp Phe Ser Phe Glu Val Ile Gly Arg Leu
260 265 270

Asp Thr Met Val Thr Ala Leu Ser Cys Cys Gln Glu Ala Tyr Gly Val
275 280 285

Ser Val Ile Val Gly Val Pro Pro Asp Ser Gln Asn Leu Ser Met Asn
290 295 300

Pro Met Leu Leu Leu Ser Gly Arg Thr Trp Lys Gly Ala Ile Phe Gly
305 310 315 320

Gly Phe Lys Ser Lys Asp Ser Val Pro Lys Leu Val Ala Asp Phe Met
325 330 335

Ala Lys Lys Phe Ala Leu Asp Pro Leu Ile Thr His Val Leu Pro Phe
340 345 350

Glu Lys Ile Asn Glu Gly Phe Asp Leu Leu Arg Ser Gly Glu Ser Ile
355 360 365

Arg Thr Ile Leu Thr Phe
370

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1128 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

```
ATG AGC ACA GCA GGA AAA GTA ATA AAA TGC AAA GCG GCT GTG CTG TGG      48
Ser Thr Ala Gly Lys Val Ile Lys Cys Lys Ala Ala Val Leu Trp
 1               5                  10                  15

GAG GAA AAG AAA CCA TTT TCC ATC GAG GAG GTG GAG GTT GCA CCC CCG      96
Glu Glu Lys Lys Pro Phe Ser Ile Glu Glu Val Glu Val Ala Pro Pro
            20                  25                  30

AAG GCC CAT GAA GTC CGT ATA AAG ATG GTG GCT ACA GGA ATT TGT CGC     144
Lys Ala His Glu Val Arg Ile Lys Met Val Ala Thr Gly Ile Cys Arg
        35                  40                  45

TCA GAT GAC CAC GTA GTT AGT GGA ACC CTT GTC ACA CCT CTT CCT GTG     192
Ser Asp Asp His Val Val Ser Gly Thr Leu Val Thr Pro Leu Pro Val
    50                  55                  60

ATC GCA GGC CAT GAG GCA GCG GGC ATT GTG GAG AGC ATT GGA GAA GGC     240
Ile Ala Gly His Glu Ala Ala Gly Ile Val Glu Ser Ile Gly Glu Gly
65                  70                  75

GTC ACT ACA GTA AGA CCA GGT GAT AAA GTC ATC CCA CTC TTT ACT CCC     288
Val Thr Thr Val Arg Pro Gly Asp Lys Val Ile Pro Leu Phe Thr Pro
 80                  85                  90                  95

CAG TGT GGA AAA TGC AGG GTT TGT AAG CAC CCT GAA GGC AAC CTC TGC     336
Gln Cys Gly Lys Cys Arg Val Cys Lys His Pro Glu Gly Asn Leu Cys
            100                 105                 110

TTG AAA AAT GAT CTG AGC ATG CCT CGG GGA ACC ATG CAG GAT GGT ACC     384
Leu Lys Asn Asp Leu Ser Met Pro Arg Gly Thr Met Gln Asp Gly Thr
        115                 120                 125

AGC AGG TTC ACC TGC AGA GGG AAG CCC ATC CAC CAC TTC CTT GGC ACC     432
Ser Arg Phe Thr Cys Arg Gly Lys Pro Ile His His Phe Leu Gly Thr
    130                 135                 140

AGC ACC TTC TCC CAG TAC ACC GTG GTG GAC GAG ATC TCA GTG GCC AAG     480
Ser Thr Phe Ser Gln Tyr Thr Val Val Asp Glu Ile Ser Val Ala Lys
145                 150                 155

ATC GAT GCG GCC TCA CCG CTG GAG AAA GTC TGT CTC ATT GGC TGT GGA     528
Ile Asp Ala Ala Ser Pro Leu Glu Lys Val Cys Leu Ile Gly Cys Gly
160                 165                 170                 175

TTT TCT ACT GGT TAT GGG TCT GCA GTC AAG GTT GCC AAG GTC ACC CAG     576
Phe Ser Thr Gly Tyr Gly Ser Ala Val Lys Val Ala Lys Val Thr Gln
            180                 185                 190

GGC TCC ACC TGT GCC GTG TTT GGC CTT GGA GGA GTG GGC CTG TCT GTT     624
Gly Ser Thr Cys Ala Val Phe Gly Leu Gly Gly Val Gly Leu Ser Val
        195                 200                 205

ATC ATG GGC TGT AAA GCA GCC GGA GCG GCC AGG ATC ATT GGG GTG GAC     672
Ile Met Gly Cys Lys Ala Ala Gly Ala Ala Arg Ile Ile Gly Val Asp
    210                 215                 220

ATC AAC AAA GAC AAG TTT GCA AAG GCC AAA GAA GTG GGT GCC ACT GAG     720
Ile Asn Lys Asp Lys Phe Ala Lys Ala Lys Glu Val Gly Ala Thr Glu
225                 230                 235

TGT GTC AAC CCT CAG GAC TAC AAG AAA CCC ATC CAG GAG GTG CTG ACA     768
Cys Val Asn Pro Gln Asp Tyr Lys Lys Pro Ile Gln Glu Val Leu Thr
240                 245                 250                 255

GAA ATG AGC AAT GGA GGT GTG GAT TTT TCC TTT GAA GTC ATT GGT CGG     816
Glu Met Ser Asn Gly Gly Val Asp Phe Ser Phe Glu Val Ile Gly Arg
            260                 265                 270

CTC GAC ACT ATG GTG ACT GCC TTG TCA TGC TGT CAA GAA GCA TAT GGT     864
Leu Asp Thr Met Val Thr Ala Leu Ser Cys Cys Gln Glu Ala Tyr Gly
        275                 280                 285

GTG AGC GTC ATT GTG GGA GTA CCT CCT GAT TCC CAA AAT CTC TCT ATG     912
Val Ser Val Ile Val Gly Val Pro Pro Asp Ser Gln Asn Leu Ser Met
    290                 295                 300

AAT CCT ATG TTG CTA CTG AGT GGA CGT ACC TGG AAA GGA GCT ATT TTT     960
Asn Pro Met Leu Leu Leu Ser Gly Arg Thr Trp Lys Gly Ala Ile Phe
305                 310                 315
```

```
GGC GGT TTT AAG AGT AAA GAT TCT GTC CCC AAA CTT GTG GCC GAT TTT      1008
Gly Gly Phe Lys Ser Lys Asp Ser Val Pro Lys Leu Val Ala Asp Phe
320                 325                 330                 335

ATG GCT AAA AAG TTT GCA CTG GAT CCT TTA ATC ACC CAT GTT TTA CCT      1056
Met Ala Lys Lys Phe Ala Leu Asp Pro Leu Ile Thr His Val Leu Pro
                340                 345                 350

TTT GAA AAA ATA AAT GAA GGA TTT GAC CTG CTT CGC TCT GGA GAG AGT      1104
Phe Glu Lys Ile Asn Glu Gly Phe Asp Leu Leu Arg Ser Gly Glu Ser
            355                 360                 365

ATC CGT ACC ATC CTG ACG TTT TGA                                      1128
Ile Arg Thr Ile Leu Thr Phe
        370
```

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 374 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

```
Ser Thr Ala Gly Lys Val Ile Lys Cys Lys Ala Ala Val Leu Trp Glu
 1               5                  10                  15

Glu Lys Lys Pro Phe Ser Ile Glu Glu Val Glu Val Ala Pro Pro Lys
            20                  25                  30

Ala His Glu Val Arg Ile Lys Met Val Ala Thr Gly Ile Cys Arg Ser
        35                  40                  45

Asp Asp His Val Val Ser Gly Thr Leu Val Thr Pro Leu Pro Val Ile
    50                  55                  60

Ala Gly His Glu Ala Ala Gly Ile Val Glu Ser Ile Gly Glu Gly Val
65                  70                  75                  80

Thr Thr Val Arg Pro Gly Asp Lys Val Ile Pro Leu Phe Thr Pro Gln
                85                  90                  95

Cys Gly Lys Cys Arg Val Cys Lys His Pro Glu Gly Asn Leu Cys Leu
            100                 105                 110

Lys Asn Asp Leu Ser Met Pro Arg Gly Thr Met Gln Asp Gly Thr Ser
        115                 120                 125

Arg Phe Thr Cys Arg Gly Lys Pro Ile His His Phe Leu Gly Thr Ser
    130                 135                 140

Thr Phe Ser Gln Tyr Thr Val Val Asp Glu Ile Ser Val Ala Lys Ile
145                 150                 155                 160

Asp Ala Ala Ser Pro Leu Glu Lys Val Cys Leu Ile Gly Cys Gly Phe
                165                 170                 175

Ser Thr Gly Tyr Gly Ser Ala Val Lys Val Ala Lys Val Thr Gln Gly
            180                 185                 190

Ser Thr Cys Ala Val Phe Gly Leu Gly Gly Val Gly Leu Ser Val Ile
        195                 200                 205

Met Gly Cys Lys Ala Ala Gly Ala Ala Arg Ile Ile Gly Val Asp Ile
    210                 215                 220

Asn Lys Asp Lys Phe Ala Lys Ala Lys Glu Val Gly Ala Thr Glu Cys
225                 230                 235                 240

Val Asn Pro Gln Asp Tyr Lys Lys Pro Ile Gln Glu Val Leu Thr Glu
                245                 250                 255

Met Ser Asn Gly Gly Val Asp Phe Ser Phe Glu Val Ile Gly Arg Leu
            260                 265                 270

Asp Thr Met Val Thr Ala Leu Ser Cys Cys Gln Glu Ala Tyr Gly Val
        275                 280                 285

Ser Val Ile Val Gly Val Pro Pro Asp Ser Gln Asn Leu Ser Met Asn
    290                 295                 300
```

```
Pro Met Leu Leu Leu Ser Gly Arg Thr Trp Lys Gly Ala Ile Phe Gly
305                 310                 315                 320

Gly Phe Lys Ser Lys Asp Ser Val Pro Lys Leu Val Ala Asp Phe Met
                325                 330                 335

Ala Lys Lys Phe Ala Leu Asp Pro Leu Ile Thr His Val Leu Pro Phe
            340                 345                 350

Glu Lys Ile Asn Glu Gly Phe Asp Leu Leu Arg Ser Gly Glu Ser Ile
        355                 360                 365

Arg Thr Ile Leu Thr Phe
    370
```

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1128 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

```
ATG AGC ACA GCA GGA AAA GTA ATA AAA TGC AAA GCG GCT GTG CTG TGG      48
    Ser Thr Ala Gly Lys Val Ile Lys Cys Lys Ala Ala Val Leu Trp
      1               5                  10                  15

GAG GAA AAG AAA CCA TTT TCC ATC GAG GAG GTG GAG GTT GCA CCC CCG      96
Glu Glu Lys Lys Pro Phe Ser Ile Glu Glu Val Glu Val Ala Pro Pro
                20                  25                  30

AAG GCC CAT GAA GTC CGT ATA AAG ATG GTG NNN ACA GGA ATT TGT CGC     144
Lys Ala His Glu Val Arg Ile Lys Met Val Ala Thr Gly Ile Cys Arg
            35                  40                  45

TCA GAT GAC CAC NNN GTT AGT GGA ACC CTT GTC ACA CCT CTT CCT GTG     192
Ser Asp Asp His Val Val Ser Gly Thr Leu Val Thr Pro Leu Pro Val
        50                  55                  60

ATC GCA GGC CAT GAG GCA GCG GGC ATT GTG GAG NNN ATT GGA GAA GGC     240
Ile Ala Gly His Glu Ala Ala Gly Ile Val Glu Xaa Ile Gly Glu Gly
    65                  70                  75

GTC ACT ACA GTA AGA CCA GGT GAT AAA GTC ATC CCA CTC TTT NNN CCC     288
Val Thr Thr Val Arg Pro Gly Asp Lys Val Ile Pro Leu Phe Xaa Pro
 80                  85                  90                  95

CAG TGT GGA AAA TGC AGG GTT TGT AAG CAC CCT GAA GGC AAC NNN TGC     336
Gln Cys Gly Lys Cys Arg Val Cys Lys His Pro Glu Gly Asn Xaa Cys
                100                 105                 110

TTG AAA AAT GAT CTG AGC ATG CCT CGG GGA ACC ATG CAG GAT GGT ACC     384
Leu Lys Asn Asp Leu Ser Met Pro Arg Gly Thr Met Gln Asp Gly Thr
            115                 120                 125

AGC AGG TTC ACC TGC AGA GGG AAG CCC ATC CAC CAC TTC CTT GGC ACC     432
Ser Arg Phe Thr Cys Arg Gly Lys Pro Ile His His Phe Leu Gly Thr
        130                 135                 140

AGC ACC TTC TCC CAG TAC ACC GTG GTG GAC GAG ATC TCA GTG GCC AAG     480
Ser Thr Phe Ser Gln Tyr Thr Val Val Asp Glu Ile Ser Val Ala Lys
    145                 150                 155

ATC GAT GCG GCC TCA CCG CTG GAG AAA GTC TGT CTC ATT GGC TGT GGA     528
Ile Asp Ala Ala Ser Pro Leu Glu Lys Val Cys Leu Ile Gly Cys Gly
160                 165                 170                 175

TTT NNN ACT GGT TAT GGG TCT GCA GTC AAG GTT GCC AAG GTC ACC CAG     576
Phe Xaa Thr Gly Tyr Gly Ser Ala Val Lys Val Ala Lys Val Thr Gln
                180                 185                 190

GGC TCC ACC TGT GCC GTG TTT GGC CTT GGA GGA GTG GGC CTG TCT GTT     624
Gly Ser Thr Cys Ala Val Phe Gly Leu Gly Gly Val Gly Leu Ser Val
            195                 200                 205

ATC ATG GGC TGT AAA GCA GCC GGA GCG GCC AGG ATC ATT GGG GTG GAC     672
Ile Met Gly Cys Lys Ala Ala Gly Ala Ala Arg Ile Ile Gly Val Asp
```

```
          210                 215                 220

ATC AAC AAA GAC AAG TTT GCA AAG GCC AAA GAA GTG GGT GCC ACT GAG      720
Ile Asn Lys Asp Lys Phe Ala Lys Ala Lys Glu Val Gly Ala Thr Glu
225                 230                 235

TGT GTC AAC CCT CAG GAC TAC AAG AAA CCC ATC CAG GAG GTG CTG ACA      768
Cys Val Asn Pro Gln Asp Tyr Lys Lys Pro Ile Gln Glu Val Leu Thr
240                 245                 250                 255

GAA NNN AGC AAT GGA GGT GTG GAT TTT TCC TTT GAA NNN ATT GGT CGG      816
Glu Xaa Ser Asn Gly Gly Val Asp Phe Ser Phe Glu Xaa Ile Gly Arg
                260                 265                 270

CTC GAC ACT ATG GTG ACT GCC TTG TCA TGC NNN CAA GAA GCA TAT GGT      864
Leu Asp Thr Met Val Thr Ala Leu Ser Cys Xaa Gln Glu Ala Tyr Gly
            275                 280                 285

GTG AGC GTC ATT NNN GGA GTA CCT CCT NNN TCC CAA AAT CTC TCT ATG      912
Val Ser Val Ile Xaa Gly Val Pro Pro Xaa Ser Gln Asn Leu Ser Met
        290                 295                 300

AAT CCT ATG TTG CTA CTG AGT GGA CGT ACC TGG AAA GGA GCT ATT TTT      960
Asn Pro Met Leu Leu Leu Ser Gly Arg Thr Trp Lys Gly Ala Ile Phe
    305                 310                 315

GGC GGT TTT AAG AGT AAA GAT TCT GTC CCC AAA CTT GTG GCC GAT TTT     1008
Gly Gly Phe Lys Ser Lys Asp Ser Val Pro Lys Leu Val Ala Asp Phe
320                 325                 330                 335

ATG GCT AAA AAG TTT GCA CTG GAT CCT TTA ATC ACC CAT GTT TTA CCT     1056
Met Ala Lys Lys Phe Ala Leu Asp Pro Leu Ile Thr His Val Leu Pro
                340                 345                 350

TTT GAA AAA ATA AAT GAA GGA TTT GAC CTG CTT CGC TCT GGA GAG AGT     1104
Phe Glu Lys Ile Asn Glu Gly Phe Asp Leu Leu Arg Ser Gly Glu Ser
            355                 360                 365

ATC CGT ACC ATC CTG ACG TTT TGA                                    1128
Ile Arg Thr Ile Leu Thr Phe
        370
```

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 374 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

```
Ser Thr Ala Gly Lys Val Ile Lys Cys Lys Ala Ala Val Leu Trp Glu
  1               5                  10                  15

Glu Lys Lys Pro Phe Ser Ile Glu Glu Val Glu Val Ala Pro Pro Lys
            20                  25                  30

Ala His Glu Val Arg Ile Lys Met Val Xaa Thr Gly Ile Cys Arg Ser
        35                  40                  45

Asp Asp His Xaa Val Ser Gly Thr Leu Val Thr Pro Leu Pro Val Ile
    50                  55                  60

Ala Gly His Glu Ala Ala Gly Ile Val Glu Xaa Ile Gly Glu Gly Val
65                  70                  75                  80

Thr Thr Val Arg Pro Gly Asp Lys Val Ile Pro Leu Phe Xaa Pro Gln
                85                  90                  95

Cys Gly Lys Cys Arg Val Cys Lys His Pro Glu Gly Asn Xaa Cys Leu
            100                 105                 110

Lys Asn Asp Leu Ser Met Pro Arg Gly Thr Met Gln Asp Gly Thr Ser
        115                 120                 125

Arg Phe Thr Cys Arg Gly Lys Pro Ile His His Phe Leu Gly Thr Ser
    130                 135                 140

Thr Phe Ser Gln Tyr Thr Val Val Asp Glu Ile Ser Val Ala Lys Ile
145                 150                 155                 160
```

Asp Ala Ala Ser Pro Leu Glu Lys Val Cys Leu Ile Gly Cys Gly Phe
165 170 175

Xaa Thr Gly Tyr Gly Ser Ala Val Lys Val Ala Lys Val Thr Gln Gly
180 185 190

Ser Thr Cys Ala Val Phe Gly Leu Gly Gly Val Gly Leu Ser Val Ile
195 200 205

Met Gly Cys Lys Ala Ala Gly Ala Ala Arg Ile Ile Gly Val Asp Ile
210 215 220

Asn Lys Asp Lys Phe Ala Lys Ala Lys Glu Val Gly Ala Thr Glu Cys
225 230 235 240

Val Asn Pro Gln Asp Tyr Lys Lys Pro Ile Gln Glu Val Leu Thr Glu
245 250 255

Xaa Ser Asn Gly Gly Val Asp Phe Ser Phe Glu Xaa Ile Gly Arg Leu
260 265 270

Asp Thr Met Val Thr Ala Leu Ser Cys Xaa Gln Glu Ala Tyr Gly Val
275 280 285

Ser Val Ile Xaa Gly Val Pro Pro Xaa Ser Gln Asn Leu Ser Met Asn
290 295 300

Pro Met Leu Leu Leu Ser Gly Arg Thr Trp Lys Gly Ala Ile Phe Gly
305 310 315 320

Gly Phe Lys Ser Lys Asp Ser Val Pro Lys Leu Val Ala Asp Phe Met
325 330 335

Ala Lys Lys Phe Ala Leu Asp Pro Leu Ile Thr His Val Leu Pro Phe
340 345 350

Glu Lys Ile Asn Glu Gly Phe Asp Leu Leu Arg Ser Gly Glu Ser Ile
355 360 365

Arg Thr Ile Leu Thr Phe
370

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 31 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

CCCCGAATTC TCAAAACGTC AGGATGGTAC G 31

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 44 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CCCCTCTAGA ATAAATGAGC ACAGCAGGAA AAGTAATAAA ATGC 44

WHAT IS CLAIMED IS:

1. A method of obtaining a nonnative protein having a thermostability that is increased over that of the native version of said protein, wherein said method comprises:

(a) obtaining in a vector a gene that encodes said native protein;

(b) mutating said vector at more than one position in said gene to produce a vector library of cells comprising mutated versions of said gene;

(c) introducing said vector library *en masse* into cells of a strain in which the majority of said mutated versions of said gene are transcribed and translated to produce a cell library;

(d) screening said cell library to identify a cell comprising a mutated version of said gene that encodes a nonnative protein having a thermostability that is increased over that of the wild-type version of said protein; and

(e) purifying said cell from said cell library.

2. The method of claim 1 which further comprises isolating from said cell in a vector said mutated version of said gene and, on said mutated version of said gene, repeating steps (b) through (e).

3. The method of claim 1 wherein said protein is an alcohol dehydrogenase.

4. The method of claim 1 wherein said protein is horse liver alcohol dehydrogenase.

5. The method of claim 1, wherein said screen is carried out in the presence of alcohol.

6. The method of claim 1, wherein said screen is carried out at an increased temperature.

7. The method of claim 1, wherein said strain is either *Escherichi coli* or *Thermus flavus*.

8. A method for selecting against growth of *Escherichi coli* recombinant cells which comprise levels

of alcohol dehydrogenase that are higher than those of wild-type *Escherichia coli* cells, wherein said method comprises growing said recombinant cells under conditions selected from the group consisting of wherein ethanol is present in a concentration of about 10%, isopropanol is present in a concentration of about 4%, and propanol is present in a concentration of about 2%, with the proviso that said wild-type cells exhibit reduced or an absence of growth under said conditions.

9. A method for selecting for growth of *Thermus flavus* recombinant cells which comprise levels of alcohol dehydrogenase that are higher than those of wild-type *Thermus flavus* cells, wherein said method comprises growing said recombinant cells under conditions selected from the group consisting of wherein ethanol is present in a concentration of about 1% in a liquid or solid medium at a pH of about 7.0, and isopropanol is present in a concentration of from about 0.5% to about 1% in a liquid or solid medium at a pH of about 7.0, with the proviso that said wild-type cells exhibit reduced or an absence of growth under said conditions.

10. A method of increasing the thermostability of horse liver alcohol dehydrogenase, which comprises introducing into a gene which encodes said alcohol dehydrogenase a mutation at a codon which codes for an amino acid residue at a position selected from the group consisting of amino acid positions 75, 94, 110, 177, 257, 268, 282, 292, and 297.

11. A method of increasing the thermostability of horse liver alcohol dehydrogenase, which comprises changing an amino acid residue at a position selected from the group consisting of amino acid positions 75, 94, 110, 177, 257, 268, 282, 292, and 297.

12. An isolated and purified nucleic acid comprising a sequence selected from the group consisting of SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, SEQ ID NO:9,

SEQ ID NO:11, SEQ ID NO:13, SEQ ID NO:15, SEQ ID NO:17, and SEQ ID NO:19.

13. An isolated and purified protein comprising a sequence selected from the group consisting of SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:16, SEQ ID NO:18, and SEQ ID NO:20.

14. A plasmid comprising the nucleic acid sequence of claim 12.

15. A plasmid selected from the group consisting of pAD7, pAD8, pAD10, pAD91, pAD92, pAD93, pAD95, pAD111, pAD113, and pTG450.

16. A vector library comprising an isolated and purified mixture of vectors comprising mutated versions of a horse liver alcohol dehydrogenase gene.

17. A host cell comprising a plasmid according to claim 14.

18. A host cell comprising a plasmid according to claim 15.

19. A host cell according to claim 17, wherein said cell is a member of the genus of *Thermus* or *Escherichia*.

20. A host cell according to claim 18, wherein said cell is strain TGF650.

21. A cell library comprising an isolated and purified mixture of cells obtained by transformation *en masse* with the vector library of claim 16.

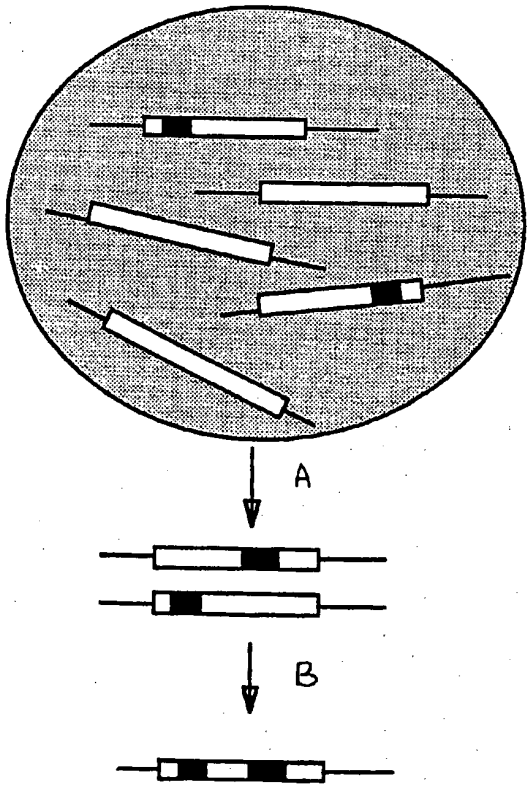

Fig 1

Fig. 2A

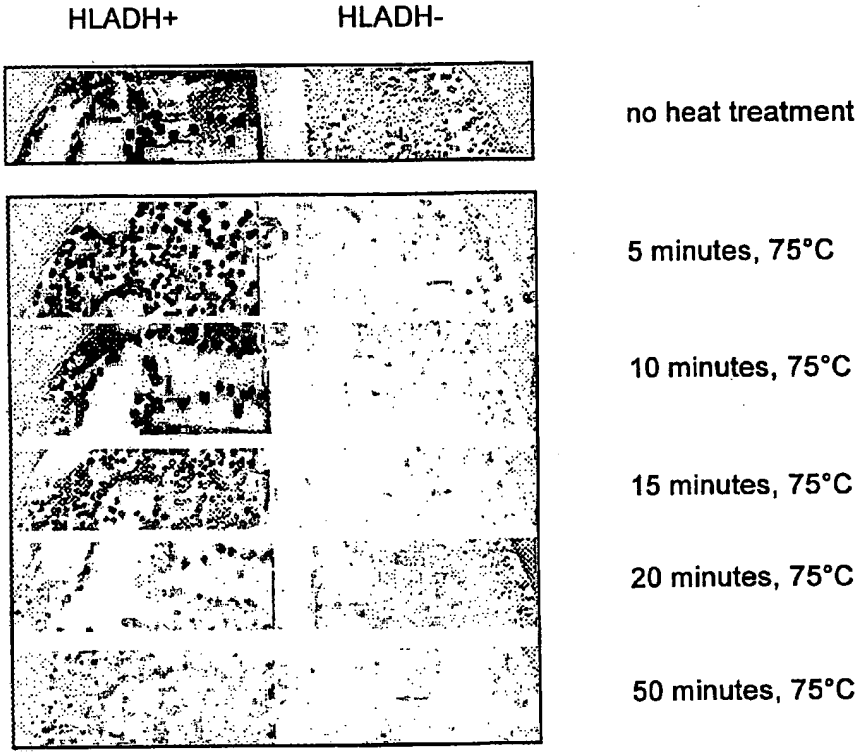
HLADH+
HLADH-
no heat treatment
5 minutes, 75°C
10 minutes, 75°C
15 minutes, 75°C
20 minutes, 75°C
50 minutes, 75°C

Fig. 2B

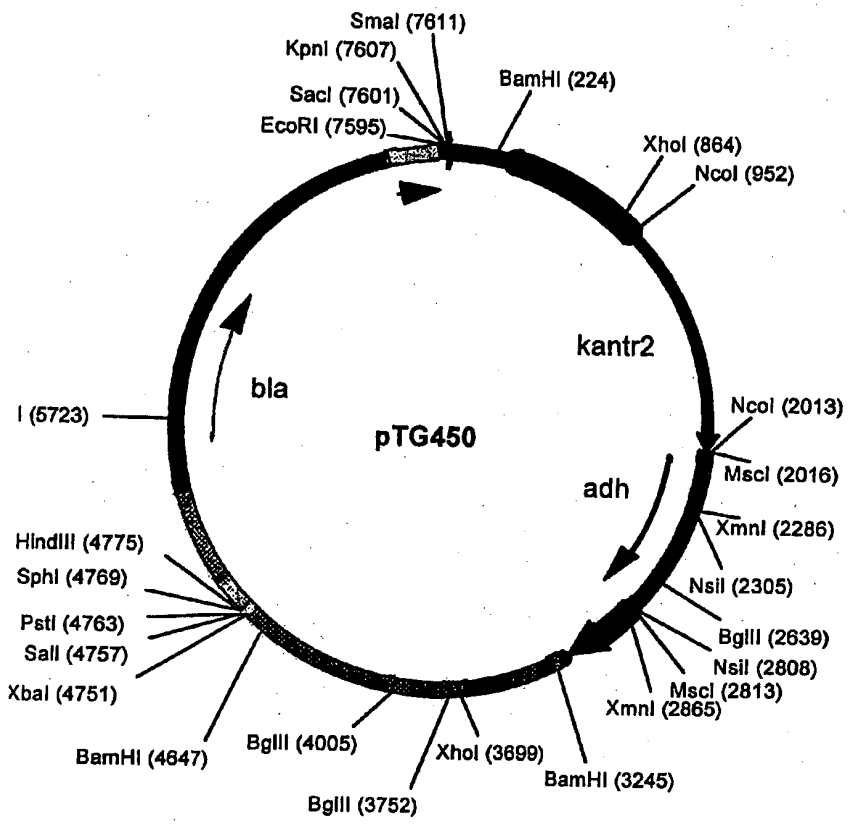
SmaI (7611)
KpnI (7607)
SacI (7601)
EcoRI (7595)
BamHI (224)
XhoI (864)
NcoI (952)
kantr2
bla
pTG450
adh
I (5723)
NcoI (2013)
MscI (2016)
XmnI (2286)
NsiI (2305)
BglII (2639)
NsiI (2808)
MscI (2813)
XmnI (2865)
HindIII (4775)
SphI (4769)
PstI (4763)
SalI (4757)
XbaI (4751)
BamHI (4647)
BglII (4005)
XhoI (3699)
BamHI (3245)
BglII (3752)

Fig. 3

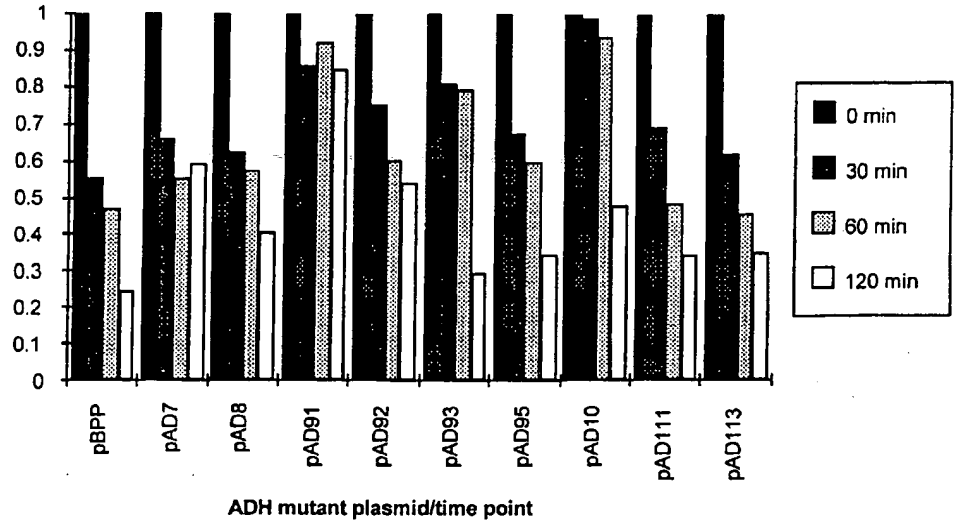
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
pBPP
pAD7
pAD8
pAD91
pAD92
pAD93
pAD95
pAD10
pAD111
pAD113
ADH mutant plasmid/time point
0 min
30 min
60 min
120 min

Fig. 4

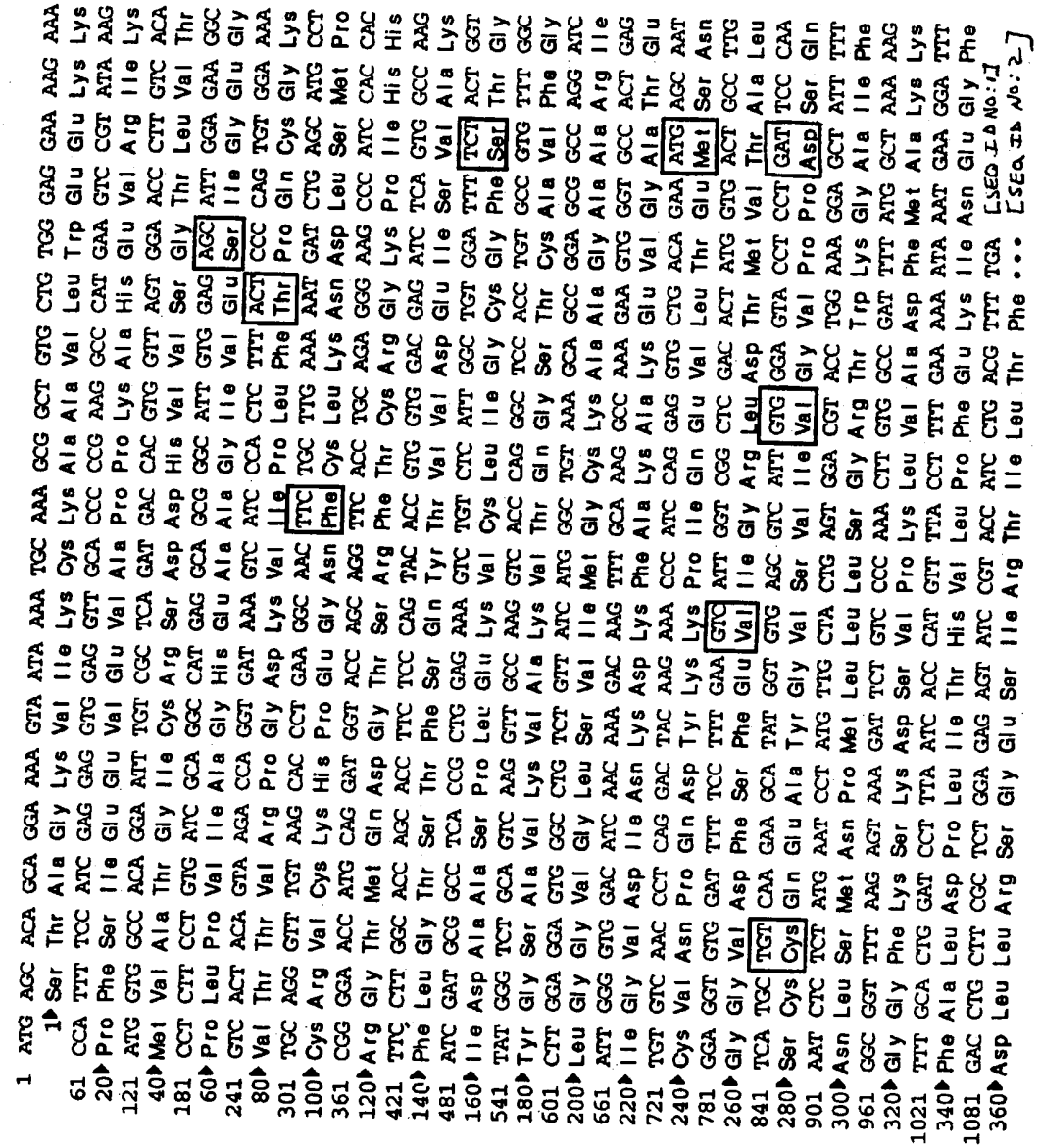

Fig. 5

INTERNATIONAL SEARCH REPORT

Inte ional Application No
PCT/US 98/09627

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/53 C12N9/04 C12N1/21

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	RELLOS P ET AL: "Polymerase chain reaction-based random mutagenesis: production and characterization of thermostable mutants of Zymomonas mobilis alcohol dehydrogenase-2." PROTEIN EXPR PURIF, JUN 1994, 5 (3) P270-7, XP002076286 UNITED STATES	1-3,5-8
Y	see page 271; figure 2 ---	4
Y	DE BOLLE X ET AL: "Identification of residues potentially involved in the interactions between subunits in yeast alcohol dehydrogenases." EUR J BIOCHEM, JUL 1 1995, 231 (1) P214-9, XP002076287 GERMANY see the whole document ---	4
	-/--	

[X] Further documents are listed in the continuation of box C. [ ] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
2 September 1998	21/09/1998

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Espen, J

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

Inte tional Application No
PCT/US 98/09627

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
X	PARK DH ET AL: "Interconversion of E and S isoenzymes of horse liver alcohol dehydrogenase. Several residues contribute indirectly to catalysis." J BIOL CHEM, MAR 15 1992, 267 (8) P5527-33, XP002076288 UNITED STATES see table 1 ---	10-14, 16-19,21
X	PARK DH ET AL: "isoenzymes of horse liver alcohol dehydrogenase active on ethanol and steroids" JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 266, 1991, pages 13296-13302, XP002076289 MD US see figure 3; tables I,VI ---	12,13
A	WEBER JM ET AL: "A chromosome integration system for stable gene transfer into Thermus flavus." BIOTECHNOLOGY (N Y), MAR 1995, 13 (3) P271-5, XP002076290 UNITED STATES -----	

Form PCT/ISA/210 (continuation of second sheet) (July 1992)